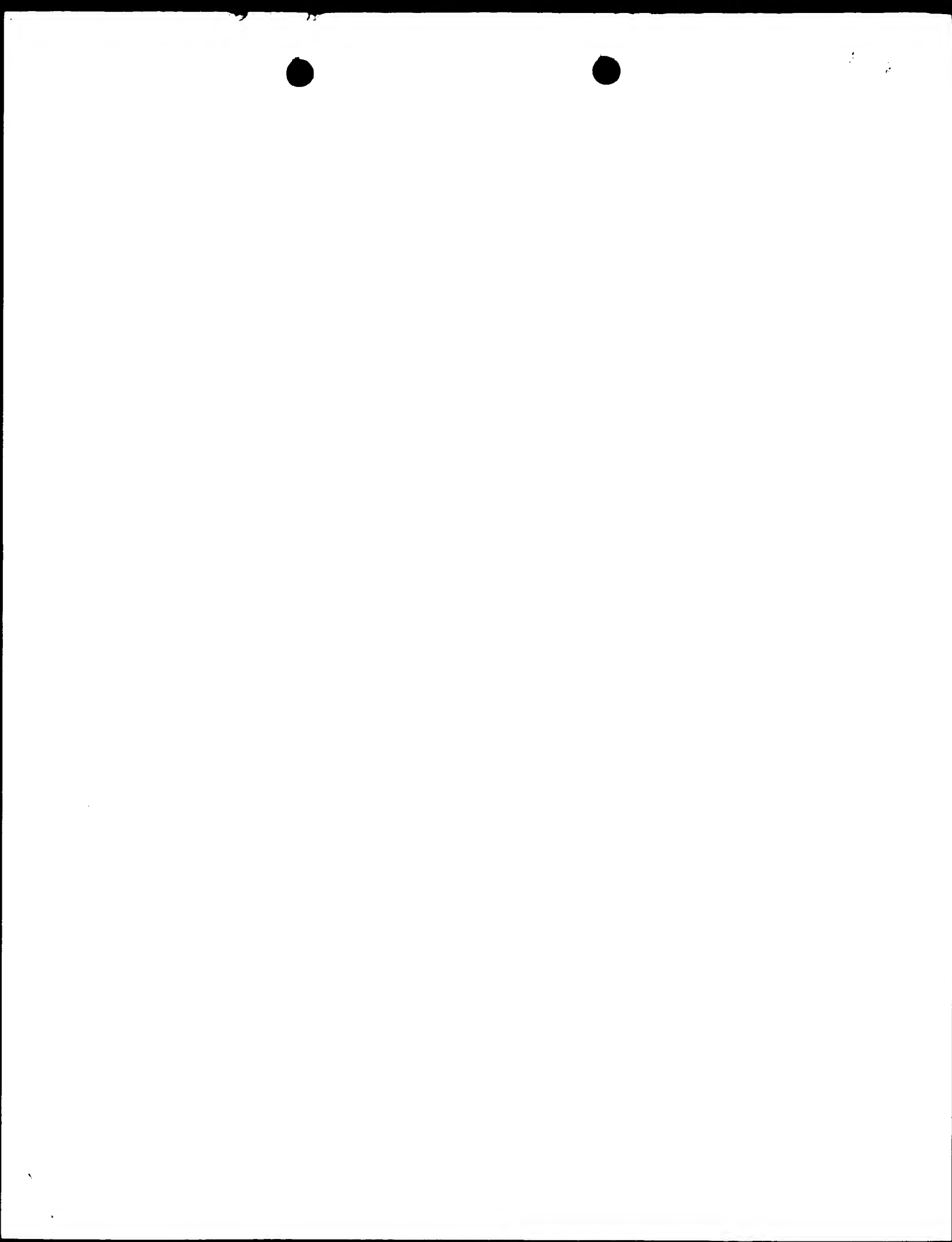


400ACACCTGGGACGACAGAGA 508
|||||:|||||



360 SerAlaPheLeuLysAlaTrpValTyrArgProGlyGluAspThrGlu1 376
 509 AGATGAGATACAGCTGAGATCCAGTCGCGACATGAGAGGTGATTCG 558
 376 uGluAsnSer.....AspLeuAspSerAlaGluG 386
 559 AAGAGCTCATGACGCAAC...ACCGGGATTAATCTGATTT..... 599
 386 LuAspThrAlaGlnThrGlyAlaThrProH1stHSerAlaPheLeuLys 402
 600 ...GGGTCCGGCGCTCAGTGAAGATTAATACCTAAAGAGAGAACCTGTA 646
 403 AlaTrpValTyrArgProGlyGluAsp...Thr.GluGluGluAsnSerA 418
 647 AAATGCCAGAGACAGTGACAGACCAACCA 678
 418 sPLeuAspSerAlaGluGluAspThrAlaGln 428

seq_name: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:US-08-244-603A-1
 seq_documentation_block:
 Sequence 1, Application US/08244603A
 Patent No. 6200768

GENERAL INFORMATION:
 APPLICANT: Mandelkow, Eva-Maria
 APPLICANT: Mandelkow, Eckhard
 APPLICANT: Lichtenberg-Kraay, Birgit
 APPLICANT: Biernat, Jacek
 APPLICANT: Drewes, Gerard
 APPLICANT: Steiner, Barbara
 TITLE OF INVENTION: No. 6200768el Tools For The Diagnosis And
 NUMBER OF SEQUENCES: 1
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
 ADDRESS: Borun
 STREET: 233 South Wacker Drive, 6300 Sears Tower
 CITY: Chicago
 STATE: Illinois
 COUNTRY: United States of America
 ZIP: 60606-6402
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Tape
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/244,603A
 FILING DATE:
 CLASSIFICATION: 435

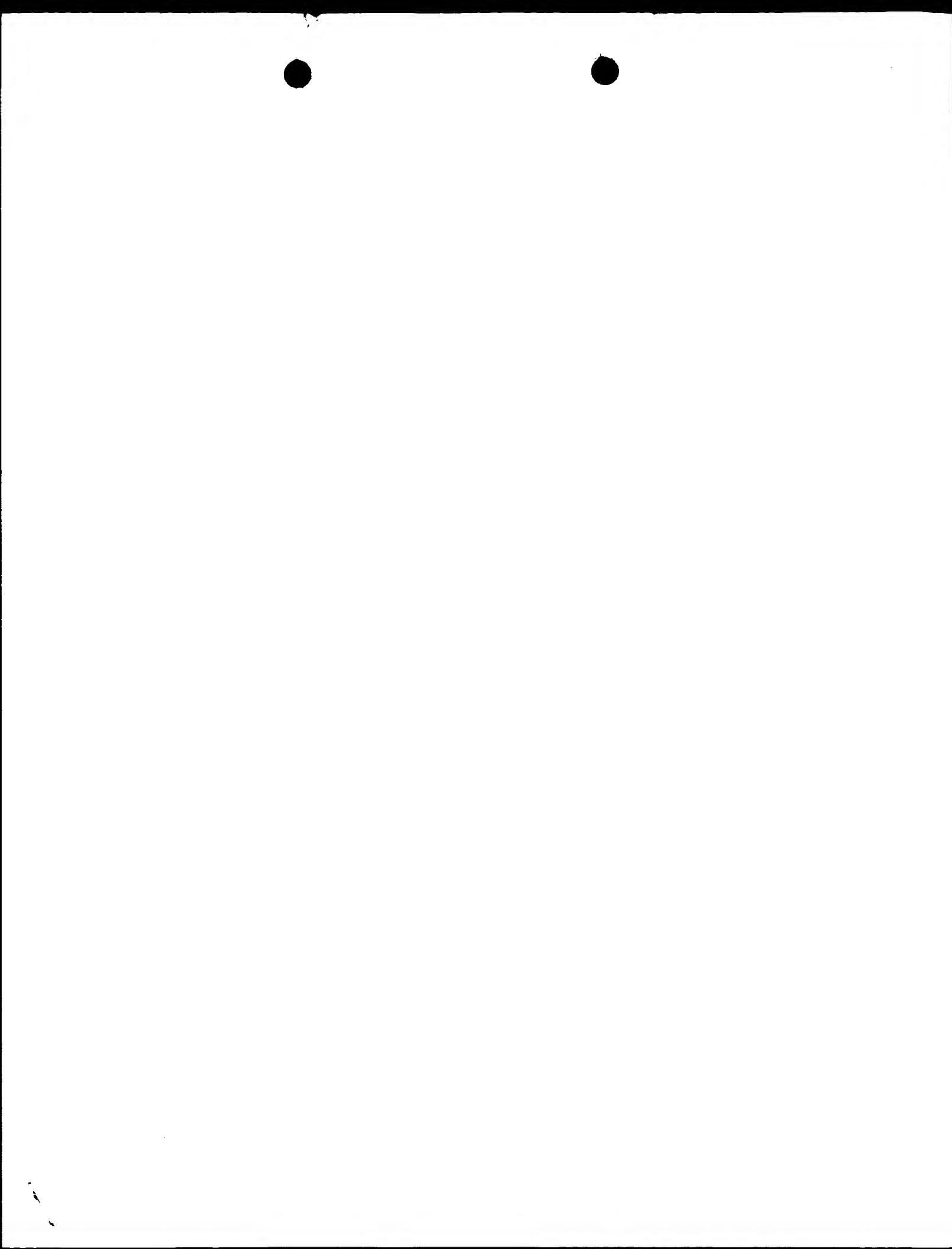
ATTORNEY/AGENT INFORMATION:
 NAME: Joseph A. Williams, Jr.
 REGISTRATION NUMBER: 38,659
 REFERENCE/DOCKET NUMBER: 28384/32778
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 312-474-6300
 TELEFAX: 312-484-0448
 TELEX: 25-3856
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 441 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 5-08-244-603A-1

Alignment scores:
 Quality: 92.00 Length: 208
 Ratio: 0.829 Gaps: 14
 Percent Similarity: 53.365 Percent Identity: 28.365

alignment block:
 US-09-651-563-808 x US-08-244-603A-1 ..
 Align seq 1/1 to: US-08-244-603A-1 from: 1 to: 441

76 GGTCCACAGCAGATCAACTGGAGTTGAGTGTGAGTGAAGA 125
 100 G1YThrThrAlaGlnGluAlaGly11eGlyAspThrProSerLeuGluAs 116
 126 GGAACACAGAGCTTCCGAGGGTGTGTGCTCACTGACTGACTGAGA 175
 116 pGluAlaAlaGly.....HisValThrGlnAlaArg 126
 176 AGGCCCTGAGAGTGTGCTCCCTCTCATGCGGTGCGACGCCCATGAGCT 225
 127 MetValSerLysSerLysAsp.....GlyThrGlySerAsp.. 138
 226 TCTGTCTGTCTACGCGCATTAATAGAGAGAGAGGCGCGAGAGTGG 275
 139AspLysLysAlaLysGlyAla 146
 276 AGGGCTCAGCGCA.....AGC 292
 146 sPGLYThrThrLysThrLysAlaThrProArgGlyAlaAlaProGlyGln 162
 293 TGGGGTGTGTGGGGGTATCCGAGTC.....CCAGAGACCC 330
 163 LysGlyGlnAlaAsnAlaThrArg11eProAlaLysThrProProAlaPr 179
 331 TGGAAACCCCGACAGAAATTTGTGACTCCCGACAGGAGACAGAGAGG 380
 179 OlyThrPro...ProSerSerGlyGluProPolysSerGlyAspArgS 195
 381 ACGCATGAGGACACACAAACACA...GACACACAGCCAGTCCA 427
 195 eRGLYTYrSerSerProGlySerProGlyThrProLysSerArgSerAr 211
 428 GGAAGCCAGTAATGAGAGAGCCCAAAAGAGAACCA.....GCAGCTG 471
 211 gThrProSerLeuProThrProProThrAlaGluProLysLysValAlaV 228
 472 AAAGTCGGATCTCAACCTGGGCGAGCAGACAGAGAAAGATGAG...GAT 518
 228 a1ValaLgThrProProLysSerProSerSerAlaLysSerArgLeuGln 244
 519 ACAGCTGATCCAGTCCAGTCCGACATGGAAGGTGATCTGCCAAGCTGCA 568
 245 ThrAlaProValProMetProAspLeuLysAsnVal...LysSerLysI 260
 569 TCAATCAACACCGGGATTAATCTGGATTTGGTCCGCGTCAAGTGG 618
 260 LeGlySerThrGluAsnLeuLysHisGlnProGlyGlyGly...LysVal 275
 619 AAGATAATACCTAAAGAG 636
 276 Gln11e11eAsnLysLys 281

seq_name: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:US-08-893-852A-1
 seq_documentation_block:
 Sequence 1, Application US/08893852A
 Patent No. 6080558
 GENERAL INFORMATION:
 APPLICANT: Bandman, Olga
 APPLICANT: Lal, Preeti
 APPLICANT: Shah, Purvi
 APPLICANT: Corley, Neil C.
 TITLE OF INVENTION: NEW HUMAN GROWTH REGULATOR PROTEIN
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Incyte Pharmaceuticals, Inc.
 STREET: 3174 Porter Drive
 CITY: Palo Alto



CC of the present invention.
XX
SQ Sequence 160 AA;

alignment_scores:
Quality: 887.00 Length: 160
Ratio: 5.544 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-651-563-808 x AAB76877 ..

Align seg 1/1 to: AAB76877 from: 1 to: 160

```

202 ATGCGGTGCGACGCCATGACCTTCTGTCTGTCACGCGCATACAGTAG 251
1 MetArgCysHisAlaHisGlyProSerCysLeuValThrAlaIleThrAr 17
252 GAGAGAGAGAGGCGCGAGAGTGGAGGCGCTCAGCGCAAGCTGGGCTGCT 301
17 ggIugIugIyglYgIProArgSerCylgIyAlaGlnAlaIalysLeuGlyCysC 34
302 GTTGGGGGATCCGAGTCCCGAAGACCTGGAGACCCCGCAGAGAGATTC 351
34 yStrIpelYlYrProSerProArgSerThrTrpAsnProAspArgPhe 50
352 TGGACTCCCGACAGCGGACGAGAGAGAGGAGCGCATGACGACACACACA 401
51 TrpThrProGlnThrIglYProGlyIugIyArgHisGluArgHisThrI 67
402 AACACGAAACACACACAGCCAGTCCCGAGGAGCCAGTATGAGAGCCCA 451
67 nThrGlnAsnHisThrAlaSerProArgSerProValMetGlnSerProL 84
452 AAAAGAGAACACGAGCTGAAAGTGGGATCTTACCTGCGGAGCAGCACA 501
84 yStrIysAsnGlnGlnLeuIysValGlyIleLeuHisLeuIleSerArg 100
502 CAGAGAGATCAGATCAGATCAGATCAGATCCGATGCGGCGCATGGAAGT 551
101 GlnIysIysIleLeuArgIleGlnLeuArgSerGlnCysAlaIleThrI 117
552 GATCTCAGAGAGTGCATCAGTCAACACCGGGGATTAATCTGGAATTGG 601
117 lIleCysIysSerCysIleSerGlnThrProGlyIleAsnLeuAspLeuG 134
602 GTTCCGCGCTCAAGGTGAAGATTAATCTTAAGAGAGACACTGTAATAATG 651
134 lYserGlyValIysValIysIleIleProIySGlnGlnHisCysIysMet 150
652 CCAAGAGCGAGTGAAGAGCAACCAAGTT 681
151 ProGlnAlaIleGlyGlnGlnProGlnVal 160
seq_name: /SIDS1/gcdata/geneseq/geneseq/AA2001.DAT: AAB76869
seq_documentation_block:
ID AAB76869 standard; Protein: 126 AA.
XX
XX AAB76869;
XX
XX 12-APR-2001 (first entry)
XX
XX Human Lung tumour protein related protein sequence SEQ ID NO:791.
XX
XX Human; lung cancer; lung tumour; lung tumour protein; gene therapy;
XX
XX lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;
XX
XX cytosolic; antisense inhibition.
XX
XX Homo sapiens.
XX
XX WO200100828-A2.

```

```

XX 04-JAN-2001.
XX
XX 30-JUN-2000; 2000MO-US18061.
XX
XX 30-JUN-1999; 99US-0346492.
XX
XX 15-OCT-1999; 99US-0419356.
XX
XX 17-DEC-1999; 99US-0466867.
XX
XX 30-DEC-1999; 99US-0476300.
XX
XX 06-MAR-2000; 2000US-0519642.
XX
XX 22-MAR-2000; 2000US-0533077.
XX
XX 10-APR-2000; 2000US-0546259.
XX
XX 27-APR-2000; 2000US-0560406.
XX
XX 05-JUN-2000; 2000US-0589184.
XX
XX (CORI-) CORIXA CORP.
XX
XX Wang T, Bangur CS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;
XX
XX Ketter MW, Mannion J;
XX
XX WPI: 2001-071488/08.
XX
XX Lung tumor-associated proteins and the nucleic acids that encode them,
XX
XX useful for preventing, diagnosing and treating lung cancer -
XX
XX Example 1; Page 413; 436pp; English.

```

The present invention describes immunogenic portions of lung tumour-associated proteins (I) and the nucleic acids (NAs) that encode them. (I) have cytosolic activity and can be used in gene therapy, antisense inhibition and in vaccines. The NAs and the lung tumour-associated proteins they encode may be used in the prevention, treatment and diagnosis of diseases associated with their inappropriate expression, especially lung cancers. For example, the NAs may be administered to treat diseases by rectifying mutations or deletions in a patient's genome that affect the activity of the protein by expressing inactive proteins or to supplement the patients own production of (I). Additionally, the NAs may be used to produce the lung-tumour associated protein, according to standard recombinant DNA methodology. Conversely, antisense NA molecules may be administered to down regulate protein expression by binding with the cells own genes and preventing their expression. The NA and complementary sequences may also be used as DNA probes in diagnostic assays to detect and quantitate the presence of similar NA sequences in samples, and hence which patients may be in need of treatment for lung cancer. The (I) may be used as antigens in the production of antibodies and in assays to identify modulators (agonists and antagonists) of the expression and activity of the protein. AAF68083 to AAF68878 and CC AAB76848 to AAB76878 represent human lung tumour protein related CC nucleotide and protein sequences which are used in the exemplification of the present invention.

XX Sequence 126 AA;

alignment_scores:
Quality: 550.00 Length: 227
Ratio: 4.583 Gaps: 2
Percent Similarity: 52.863 Percent Identity: 52.423

alignment_block:
US-09-651-563-808 x AAB76869 ..

Align seg 1/1 to: AAB76869 from: 1 to: 126

```

2 CGGCGAGAGTGTGAGCGCGGACTCGGGTCCGTGAGTGTGATTTTC 51
1 ArgArgSerCysGlnProAlaIleThrArgValProGlnValIleThrIleuSe 23
52 TCCGCTACTGAGACAGCGCGGTAAGTCCACAGCAGATCCACACTGGAG 101
23 rProIleuIleArgHisGly..... 29
102 TTGAAGTGTGAGTGAAGTGAAGAGAACCAAGCAGAGCTTCGAGAGGTTG 151

```



```

23 ..... 23
252 GGAGGAGGAGGCGCGAGAGTGGAGGGGCTCAGAGCGAAGCTGGGTCT 301
23 ..... 23
302 GTTGGGGGTATCGAGTCCGAGAGCAGCTTGAACCCCGACAGAGATTC 351
23 ..... 23
352 TGGACTCCCGCAGCGGACAGAGAGGAGCGGCGATGAGCGACACACACA 401
24 .....Gly..HisThrG1 27
402 AACAGAGAACACACAGCCAGTCCCGAGAGCCGAGTAATGAGAGCCCA 451
|||||
27 nThrglnasnHisThrAlaSerProAlaSerProValMetGlnSerProl 44
452 AAAAGAGAGACCCAGCAGCTGAAAGTCCGGATCTTACACCTGGGACAGACA 501
|||||
44 ySLySLySAsnglnGlnLeuLySValGlylleuHisLeuGlySerArg 60
502 CAGAGAGAGATCAGGATACAGCTGAGATCCAGTC..... 536
|||||
61 GlnySLySLyIleArgIleGlnInLeuArgSerGlnValLeuGlyArgGlu 77
537 CGCGACATGAGAGGTGATCTGCAGAGCTGCATCAGTCAACACCGGGG 585
|||||
77 TArgSPmetGlnGlyAspLeuGlnGlnLeuHisGlnSerAsnThrGly 94
586 ATAAATCTGATTTGGGTCCGGCGGTCAAGGTCAAGTCAATATACC 629
|||||
94 sPLySerGlyPhGlyPheArgArgGlnGlyLeuAspAsnThr 108

seq_name: /STD1/gcgdata/geneseq/geneseq/AA2001.DAT:AAAB76870
seq_documentation_block:
ID AAB76870 standard; Protein: 108 AA.
XX
AC AAB76870;
XX
DT 12-APR-2001 (first entry)
XX
DE Human lung tumour protein related protein sequence SEQ ID NO:793.
XX
KW Human; lung cancer; lung tumour; lung tumour protein; gene therapy;
KW lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;
KW cytosolic; antisense inhibition.
XX
OS Homo sapiens.
XX
PN WO200100828-A2.
XX
PD 04-JAN-2001.
XX
PF 30-JUN-2000; 2000MO-US18061.
XX
PR 30-JUN-1999; 9905-0346492.
PR 15-OCT-1999; 99US-0419356.
PR 17-DEC-1999; 99US-0466667.
PR 30-DEC-1999; 99US-0476300.
PR 06-MAR-1999; 99US-0519642.
PR 22-MAR-2000; 2000US-0533077.
PR 10-APR-2000; 2000US-0546259.
PR 27-APR-2000; 2000US-0560406.
PR 05-JUN-2000; 2000US-0589184.
XX
PA (CORI-) CORIXA CORP.
XX
PI Wang T, Bangur CS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;
PI Retter MW, Mannion J;
XX

```

DR WPI: 2001-071488/08.
XX
XX Lung tumor-associated proteins and the nucleic acids that encode them
PT useful for preventing, diagnosing and treating lung cancer -
XX
XX Example 1; Page 413-414; 436pp; English.
PS

The present invention describes immunogenic portions of lung tumour-associated proteins (I) and the nucleic acids (NAs) that encode them. (I) have cytosolic activity and can be used in gene therapy, antisense inhibition and in vaccines. The NAs and the lung tumour-associated proteins they encode may be used in the prevention, treatment and diagnosis of diseases associated with their inappropriate expression, especially lung cancers. For example, the NAs may be administered to treat diseases by rectifying mutations or deletions in a patient's genome that affect the activity of the protein by expressing inactive proteins or to supplement the patients own production of (I). Additionally, the NAs may be used to produce the lung tumour associated protein, according to standard recombinant DNA methodology. Conversely, antisense NA molecules may be administered to down regulate protein expression by binding with the cells own genes and preventing their expression. The NA and complementary sequences may also be used as DNA probes in diagnostic assays to detect and quantitate the presence of similar NA sequences in samples, and hence which patients may be in need of treatment for lung cancer. The (I) may be used as antigens in the production of antibodies and in assays to identify modulators (agonists and antagonists) of the expression and activity of the protein. AAF6083 to AAF68878 and AAF76848 to AAF67878 represent human lung tumour protein related nucleotide and protein sequences which are used in the exemplification of the present invention.

	alignment_scores:	
Quality:	423.50	Length: 219
Ratio:	4.152	Gaps: 3
Percent Similarity:	47.442	Percent Identity: 46.977

```
alignment_block:
```

US-09-651-563-808 x AAB76870

Align seg 1/1 to: AAB76870 from: 1 to: 108

```

2  GGGGGAGGTGTGACCGGACCGGACCTGGGTCTCGAGCTGATCTTTC  51
   |||||
1  ATGATgSeCySgLProlAtHrATgAlPrGluIvaITrPileLeSe  17
52  TCCGCTACTGAGACACGGCGGGTAGTGCACAGGCAGATCCAACTGGAG  101
   |||||
17  rProLeuEuArHtSgLy.....  23
102  TTGAAGTGTGAGTGTGAGAGTGAAGGAACACAGAGCTTCCGGAGGGTTG  151
23  .....  23
152  TGTGTCACGTACTCAGAGTGAAGAGCCCTCGAAGTGTGTCCTCTC  201
23  .....  23
202  ATGCGGTGCCACGCCCATGGAACTTTGTCTGTCGTCACGCCATTAAGTAG  251
23  .....  23
252  GGAGGAAGAGAGGCCGAGAGAGTGGAGGGGCTCAGGCCAAGCTGGGGTGT  301
23  .....  23
302  GTTGGGGGATATCCAGATCCCAAGACACTGGAACCCGACAGAAATTC  351
23  .....  23
352  TGGACTCCCCAGACGGGACACAGAGAGGAGGACGCGATGAGCGACACACACA  401

```



```

24 .....::: ||||| 27
402 AACGAGAACACACAGCCAGTCCAGAGCCAGCAATGAGAGAGCCCA 451
|||||
27 nhrhglnasnhrsthralsrProhrserProvalmetgluserProL 44
|||||
452 AAAAGAAGAACGACGAGCTGAAAGTGGAGCTCTACCTGGGACGACGA 501
|||||
44 ystysrsgnsnglnleuysvalgIlylleuhtstleuglyserarg 60
|||||
502 CAGNAAAGATCAGATACGCTGAGATCCCACTG..... 536
|||||
61 GlnLysLysIleArgIleGlnLeuAgsSerGlnValLeuGlyArgGln 77
|||||
537 .CGGACATGAGAGATGATCTGCAAGAGCTGCATGCTCAACACCGGG 585
|||||
77 targaAspmetgluglyAspLeuclncluleuhtstglnserAserThrgIYA 94
|||||
586 ATTAATCTGGATTGGTTCCGGCTCAAGGTGAAGATATAC 629
|||||
94 splysserGlyPheGlyPheArgArgGlnGlyGlnspasnThr 108
|||||
seq_name: /SIDSI/gcgdata/geneseq/geneseq/AA2001.DAT: AAB60500
seq_documentation_block:
ID AAB60500 standard; Protein: 111 AA.
AC AAB60500;
XX
XX 24-APR-2001 (first entry)
XX
XX Human cell cycle and proliferation protein CCYPR-48. SEQ ID NO:48.
XX
XX Cell cycle and proliferation protein: CCYPR: human; agonist:
XX antagonist; gene therapy; detection; gene therapy;
XX transgenic animal disease model; immune disorder;
XX developmental disorder; cell signalling disorder;
XX cell proliferative disorder; cancer; tumour; anaemia; epilepsy;
XX arteriosclerosis; asthma; allergy; diabetes mellitus;
XX menstrual cycle disorder; bacterial infection.
XX
XX Homo sapiens.
XX
XX WO200107471-A2.
XX
XX 01-FEB-2001.
XX
XX 21-JUL-2000; 2000WO-US19948.
XX
XX 21-JUL-1999; 99US-0145075.
XX 08-SEP-1999; 99US-0153129.
XX 10-NOV-1999; 99US-0164647.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Hillman JL, Lai P, Tang YT, Yue H, Au-Young J, Bandman O;
XX Azimzal Y, Yang J, Lu DAM, Baughn MR, Patterson C, Shah P;
XX WPI. 2001-112727/12.
XX N-PSDB; AAF59637.
XX
XX Human cell cycle and proliferation proteins and polynucleotides are
XX used to treat, diagnose and prevent immune, developmental and cell
XX signalling disorders and cell proliferative disorders including cancer -
XX
XX Claim 1: Page 156; 205pp; English.
XX
XX Sequences AAB60453-AAB60506 represent 54 human cell cycle and
XX proliferation proteins (CCYPR), which are encoded by AAF59590-AAF59643.
XX CCYPR and agonists of CCYPR are used to treat diseases or conditions
XX associated with decreased expression of functional CCYPR, while CCYPR
XX antagonists are used to treat diseases or conditions associated with

```

```

CC overexpression of functional CCYPR. Monoclonal or polyclonal antibodies
CC to CCYPR may be used in enzyme-linked immunosorbent assays (ELISA) or
CC radioimmunoassays to detect CCYPR. CCYPR itself may be used to detect
CC compounds e.g., antibodies, oligonucleotides and proteins (receptors)
CC that specifically bind to CCYPR, and in drug screening methods to
CC identify compounds that modulate the activity of CCYPR. CCYPR
CC nucleotides can be used to generate transgenic animal models of human
CC disease, and can be used in gene therapy in target cells with genetic
CC abnormalities with respect to the expression of CCYPR for the
CC treatment or prevention of a disorder associated with CCYPR.
CC diseases which can be diagnosed, treated and prevented using CCYPR
CC proteins, nucleic acids, agonists or antagonists include immune,
CC developmental and cell signalling disorders, and cell proliferative
CC disorders including cancer. Specific examples of these disorders
CC include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies,
CC diabetes mellitus, disorders of the menstrual cycle and infections
CC caused by bacteria.
XX
XX Sequence 111 AA:
SQ
alignment_scores:
Quality: 233.00 Length: 85
Ratio: 3.426 Gaps: 1
Percent Similarity: 80.000 Percent Identity: 56.471
alignment_block:
US-09-651-563-808 x AAB60500 ..
Align seg 1/1 to: AAB60500 from: 1 to: 111
429 GAGCCCACTAATGAGAGCCGCCAAAAGAGACCGACGCTGAAGTGC 478
|||||
28 GluProthrAspGluGluProGlyGlnGluProThrLysSerAr 44
|||||
479 GGATCTTACCTGGGCGACGACGAGAGATCAGATCAGTCTGGA 528
|||||
44 GasnProthrProAspGlnLysArgGluAspAspGlnLysValaLaGluI 61
|||||
529 TCCCGTCGCGGACATGAGAGTGTGTCAGAGAGCTGCATCAGTCAAC 578
|||||
61 LeglnValProAspLeuclnLalaAspLeuGlnLeuLysCglnThrLys 77
|||||
579 ACCGGGGATTAATCTGATTGGTTCCGGCTCAAGGTGAAGATATAC 628
|||||
78 ThrGlyAsp.GlyCysGlnGlyGlyThrAspValLysGlyLysIleLeuP 94
|||||
629 CTAAAGAGGACACTGTAAATGCCAGAGAGGTGAAGAGCAACCA 678
|||||
94 TolysAlaGlnHisPheLysMetProGlnAlaGlyGlnGlyLysSerGln 110
|||||
679 GTT 681
|||
111 Val 111
seq_name: /SIDSI/gcgdata/geneseq/geneseq/AA2000.DAT: AAY83158
seq_documentation_block:
ID AAY83158 standard; Protein: 102 AA.
XX
XX AAY83158;
XX
XX 24-JUL-2000 (first entry)
XX
XX PAGE-4 polypeptide.
XX
XX PAGE-4; MAGE; GAGE; reproduction; tests; prostate; fallopian tube;
XX uterus; placenta; cancer; major histocompatibility complex; MHC; CTL;
XX cytotoxic T lymphocyte; immune response; antibody; drug delivery;
XX
XX Homo sapiens.
XX

```


XX 09-MAR-2000; 2000WO-US06112.
 PE 11-MAR-1999; 9905-0266513.
 PR 18-AUG-1999; 9905-0149485.
 XX
 PA (GENE-) GENESIS RES & DEV CORP LTD.
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 XX
 PI Wood M, McGrath A, Shenk MA, Glenn M;
 XX
 DR WPI: 2000-579369/54.
 XX
 PT New isolated polynucleotide encoding a plant transcription factor for
 PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
 PT having modified gene expression or modified activity of a polypeptide
 PT
 PS
 XX
 PS Claim 8; Page 389; 747pp; English.

XX The present invention relates to novel plant transcription factors from
 CC Eucalyptus grandis or Pinus radiata. The present sequence is one such
 CC transcription factor. The transcription factor may be used to produce a
 CC plant having modified gene expression such as a woody plant e.g. a
 CC eucalyptus, pine, acacia, poplar, sweetgum, teak, or mahogany species or
 CC to modify the activity of a polypeptide in a plant. The transcription
 CC factors of the present invention are members from the following families
 CC of regulatory proteins: bZIP, bZIP family of G-box binding factors, basic
 CC helix-loop-helix zipper, homeotic/homeodomain/homeobox/MAUS, homeodomain
 CC Cys2His2, CCAR box elements and MYB.
 CC
 SO Sequence 396 AA;

alignment_scores:
 Quality: 103.50 Length: 226
 Ratio: 0.855 Gaps: 13
 Percent Similarity: 53.540 Percent Identity: 29.646

alignment_block:
 US-09-651-563-808 x AAB33006 ..

Align seg 1/1 to: AAB33006 from: 1 to: 396

```

63 GACAGGGGGGTGAGTCCAGCAGCATCCAACTGGGAGTTGAATGCA 112
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
132 AsPThrGlyArgGlyArgGlyArgGlyArgGlyArgGlyArgGly 146
113 GTGAGAGTGAAGAGAGACAGCAGGCTTCGAGAGGTTGTGTGCTAGT 162
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
146 PThrGlyArgGlyArgGlyArgGlyArgGlyArgGlyArgGlyArg 156
163 ACTCAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
156 SerGlyGlnAspAspAlaAlaGlyGlyGlyGlnSerArgGlyArg 172
201 CATGGGGTCCAGCCCATGACCTTCTGTCTCTGTCACGGCCATAACTA 250
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
173 ArgGlyArgGlyArgGlyArgGlyArgGlyArgGlyArgGlyArg 181
251 GGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 eGlyGlyArgGlyArgGlyArgGlyArgGlyArgGlyArgGlyArg 195
301 TGTGGGGTATCCAGTCCAGAGACACTGAGACCCGACAGAGAGATT 350
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
196 ..TyrGlyTyrAspAlaAsnGlyGlnAspArgProPArgGlnGln 211
351 CTGACCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 400
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
212 GlnAspThr.....AsnThrPheThrGlyThr.....AspAsnTr 223
  
```

```

401 AAACAGAGACCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 450
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
223 AsPThrProGlnValSerValValAspGlnValValAsnValGlnProG 240
451 AAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 500
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
240 LngLysLysProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 252
501 ACAGAGAGAGATCAGATCAGATCAGATCAGATCAGATCAGATCAG 550
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
253 ThrSerGlnAsnLysAspAsnLysGlnGlnGlnGlnGlnGlnGln 269
551 TGATCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
269 AsPGLTyrGlnLysLysLysLysLysLysLysLysLysLysLys 285
601 GGTTCGGCGCTCAAGTGAAGATTAATCACTTAAGAGAGAGAGAG 650
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
286 LeuLysAsnAlaGlnArgLysValIleLeuAspArgAspPheGln 302
651 GCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 688
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
302 eGlnLeuValAspLysLysAsnAspGlyTllePheIleLysLeuAsn 318
689 AGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 710
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
318 GlnLysGlnArgGlnArgLys 325
  
```

seq_name: /SIDSI/gcgdata/geneseq/geneseq/AA1998.DAT:AAW79958

seq_documentation_block:
 ID AAW79958 strand: Protein; 578 AA.

```

XX AC AAW79958;
XX DT 02-FEB-1999 (first entry)
XX DE Human progression elevated gene-2 (PEG-3) protein.
XX KW Progression elevated gene-3; PEG-3; human; tumour progression;
XX KW DNA repair; tumorigenesis; angiogenesis; metastasis; melanoma;
XX KW brain cancer; cervix cancer; prostate cancer; lung cancer;
XX KW colorectal cancer; neuroblastoma; glioblastoma; diagnosis; therapy.
XX OS Homo sapiens.
XX PN WO9842315-A1.
XX PD 01-OCT-1998.
XX PF 20-MAR-1998; 98WO-US05793.
XX PR 21-MAR-1997; 97US-0821818.
XX PA (GENO-) GENQUEST INC.
XX PA (UYCO) UNIV COLOMBIA NEW YORK.
XX PI Fisher PB, Goldstein NI, Su Z, Zhang N;
XX DR WPI: 1998-557025/47.
XX DR N-PSDB: AAV65765.
XX PT New isolated Progression Elevated Gene-3 - used to develop products
XX PT for e.g. modulating DNA damage and repair pathways, cancer
XX PT progression or oncogene mediated transformation and angiogenesis.
XX PS Claim 17; Fig 13A-C; 225pp; English.
XX CC This polypeptide comprises human progression elevated gene-3 (PEG-3)
XX CC protein. The amino acid sequence was deduced from an isolated cDNA
XX CC clone (see AAV65765). PEG-3 expression (1.5 and 2.8 kb transcripts)
XX CC is elevated in most human tumour cell lines. The amount of PEG-3
XX CC in a cell is also an indicator of DNA damage in that cell. Cells
  
```

in which a reporter gene in under control of the promoter or regulatory region (see AAV65766) of the rat PEG-3 gene can be used in methods for identifying agents that modulate PEG-3 expression or the ability of PEG-3 to induce progression, or for determining whether an agent is capable of inhibiting DNA damage and repair pathways, cancer progression or oncogene-mediated transformation. Compounds that induce DNA damage or which regulate angiogenesis can also be identified using such cells. Transgenic animals and vaccines comprising PEG-3 polypeptides and an immune response enhancer are also claimed. Antibodies (especially monoclonal) to the PEG-3 may be used to determine tumour progression of a cell. Cells can be protected from chemotherapeutic damage by inhibiting or eliminating the expression of PEG-3 in the cells. The methods can be applied to a progression phenotype comprising anchorage-independent growth, tumorigenesis, angiogenesis or metastasis, to melanoma, brain, cervical, prostate, lung or colorectal cancer, neuroblastoma or glioblastoma (all claimed).

Sequence 578 AA:

alignment_scores:

Quality: 103.00 Length: 179
Ratio: 1.198 Gaps: 11
Percent Similarity: 48.045 Percent Identity: 27.374

alignment_block:

US-09-651-563-808 x AAW79958 ..

Align seg 1/1 to: AAW79958 from: 1 to: 578

```

112 AGTGAAGTGAAGAGACAGGAGGCTCCGAGGCTTGTGTCAGT 161
    : : : : : : : : : : : : : : : : : : : : : : : :
220 ThrAspAsnIysAlaIupProSerGlySerHisSerAlaPheThrIu 236
    : : : : : : : : : : : : : : : : : : : : : : : :
162 GACTCAGAGTGAAGAGGAGCCCTGAGTCGTCCTCATCGCGTCC 211
    : : : : : : : : : : : : : : : : : : : : : : : :
236 rHsIstHrArgIuArProIySgIngluIuThrlYsProIuGlnH 253
    : : : : : : : : : : : : : : : : : : : : : : : :
212 AGCCCATGAGACCTTCTTGTCTGTCAGCGCCATAAGGAGAGGA 261
    : : : : : : : : : : : : : : : : : : : : : : : :
253 IsArGAlaIyGInSerHisProIySgInHsAlaIuSerGluIu 269
    : : : : : : : : : : : : : : : : : : : : : : : :
262 GGGCCGAGG.....AGTGAAGGCTCAGGCGCAAGTCGGG 299
    : : : : : : : : : : : : : : : : : : : : : : : :
270 GLyProIuThrSerValCysSerGlySerAlaPheIuIyS..... 283
    : : : : : : : : : : : : : : : : : : : : : : : :
300 CTGTGGGGGTATCCGAGTCCGAGAACCTGTGAACCCCGACAGAGAT 349
    : : : : : : : : : : : : : : : : : : : : : : : :
284 .AlaTrpValTyr.Arg.....ProIyGluSprIngluIu 295
    : : : : : : : : : : : : : : : : : : : : : : : :
350 TCTGACTCCCA.....GACGGAGCAGAGAGGAGCGCATGAGCGA 393
    : : : : : : : : : : : : : : : : : : : : : : : :
296 Glu.GluAspProAspleuAspSerAlaIuIuAspThrAla..... 309
    : : : : : : : : : : : : : : : : : : : : : : : :
394 CACACACAAACACAGAACACACAGC..... 419
    : : : : : : : : : : : : : : : : : : : : : : : :
310 .HsIstHrCysThrlThrProHisThrSerAlaPheIuYsAlaTrpVal 325
    : : : : : : : : : : : : : : : : : : : : : : : :
420 .....CAGTCCGAGGAGCCGATATGAGAG.....CCCCA 451
    : : : : : : : : : : : : : : : : : : : : : : : :
326 TyrArgProIyGluSprIngluIuAspAspIyAspTrpAspSe 342
    : : : : : : : : : : : : : : : : : : : : : : : :
452 AAAGAAGAACACAGCTGAAGT...CGGATCCT..... 485
    : : : : : : : : : : : : : : : : : : : : : : : :
342 rAlaIuGluAlaIuSerGlySerHisSerAlaPheThrIuIu 359
    : : : : : : : : : : : : : : : : : : : : : : : :
486 .....ACACTGGGACACAGACAGAGAGATCAG 515
    : : : : : : : : : : : : : : : : : : : : : : : :
359 heIuYsAlaTrpValTyrArgProIyGluSprIngluIuAsp 375
    : : : : : : : : : : : : : : : : : : : : : : : :
516 GATACAGTGAATCCAGTGGCGGACATGGA 548
    : : : : : : : : : : : : : : : : : : : : : : : :

```

376 AspSerGluAsnValAlaProValAspSerGlu 386

seq_name: /SIDSI/gcdata/geneseq/geneseq/AA1999.DAT:AAV41104

seq_documentation_block:

ID AAV41104 standard; Protein; 578 AA.

AAV41104:

17-JAN-2000 (first entry)

Human progression elevated gene-3 (PEG-3) amino acid sequence.

Progression-elevated gene-3; PEG-3 gene; regulatory region; cell death;

cell growth; cancer; thymidine kinase; gancyclovir; acyclovir; human.

Homo sapiens.

MOJ949898-A1.

07-OCT-1999.

31-MAR-1999; 99MO-US07199.

31-MAR-1998; 98US-0052753.

(UYCO) UNIV COLUMBIA NEW YORK.

Fisher PB:

WPI: 1999-591184/50.

N-PSDB; AA223029.

Novel vectors useful for studying the progression of cancer -
Disclosure; Fig 13A-C; 251pp; English.

The invention relates to an inducible progression-elevated gene-3 (PEG-3 gene) regulatory region functionally linked to a gene encoding a product that causes or may be induced to cause the death or inhibition of cancer cell growth. A vector of the invention which contains a gene encoding thymidine kinase or a product which causes the cell to express a specific antigen can be administered along with gancyclovir or acyclovir, or an antibody or fragment to the antigen, respectively, to treat cancer in a subject. The PEG-3 gene is useful for generating new cloning and expression vectors, transfected cells, and for developing methods for cultured growth of such cells. The PEG-3 polynucleotide is also useful as a source of primers and probes to study the progression of cancer, and to detect the presence of the gene. The present sequence represents the amino acid sequence of the human PEG-3 protein.

Sequence 578 AA:

alignment_scores:

Quality: 103.00 Length: 179
Ratio: 1.198 Gaps: 11
Percent Similarity: 48.045 Percent Identity: 27.374

alignment_block:

US-09-651-563-808 x AAV41104 ..

Align seg 1/1 to: AAV41104 from: 1 to: 578

```

112 AGTGAAGTGAAGAGACAGGAGGCTCCGAGGCTTGTGTCAGT 161
    : : : : : : : : : : : : : : : : : : : : : : : :
220 ThrAspAsnIysAlaIupProSerGlySerHisSerAlaPheThrIu 236
    : : : : : : : : : : : : : : : : : : : : : : : :
162 GACTCAGAGTGAAGAGGCGCCGTGAAGTCGTCCTCATCGCGTCC 211
    : : : : : : : : : : : : : : : : : : : : : : : :
236 rHsIstHrArgIuArProIySgIngluIuThrlYsProIuGlnH 253
    : : : : : : : : : : : : : : : : : : : : : : : :
212 AGCCCATGAGACCTTCTTGTCTGTCAGGCGCATATGAGAGGA 261
    : : : : : : : : : : : : : : : : : : : : : : : :

```

```

253 1sArgAlaIylGlnSerHisProCysGlnAsnAlaGlnSerIuGluGly 269
262 GGGCCGAGG.....AGTGGAGGGGCTCAGGCGCAAGCTGGGGT 299
270 GlyProGluThrSerValCysSerGlySerAlaPheLeuLys..... 283
300 CTGTGGGGGTATCCGAGTCCAGAAACACCTGGAAACCCGACAGAGAT 349
284 .AlaTrpValI Tyr .Arg.....ProGlyGluAspThrGluGlu 295
350 TCTGGACTCCCA.....GACGGACACGAGAGAGGCGCATGAGGA 393
296 Glu.GluAspProAspLeuAspSerAlaGluGluAspThrAla..... 309
394 CACACACAAACACAGAACCCACACAGC..... 419
310 ..HisThrCysThrThrProHisThrSerAlaPheLeuLysAlaTrpVal 325
420 .....CAGTCCGAGAGCGCCAGTATGAGAG...CCCA 451
326 TyrArgProGlyGluAspThrGluGluAspGlyAspTrpAspSe 342
452 AAAAGAGAACACGACGCTGAAGT...CGGATCCT..... 485
342 rAlaGluGluAspAlaAlaGlnSerCysThrThrProHisThrSerAlaP 359
486 .....ACACCTGGGACGACAGACAGACAGAGATGAG 515
359 heLeuLysAlaTrpValI TyrArgProGlyGluAspThrGluGluAsp 375
516 GATACAGCTGAGATCCAGTCCGCGCATGGAA 548
376 AspSerGluAsnValAlaAlaProValAspSerGlu 386
seq_name: /SIDS1/gcgdata/geneseq/geneseq/AA1999.DAT:AA06514

```

```

seq_documentation_block:
ID  AA06514 standard; Protein; 578 AA.
XX
AC  AA06514;
XX
DT  08-OCT-1999 (first entry)
XX
DE  Human progression elevated protein-3 (PEG-3).
XX
KM  Progression elevated protein-3; PEG-3; human; angiogenesis; cancer;
XX  therapy; vaccine.
XX
OS  Homo sapiens.
XX
PN  W0937776-A1.
XX
PD  29-JUL-1999.
XX
PE  26-JAN-1999; 99MO-US01623.
XX
PR  29-JAN-1998; 98US-0072941.
XX  26-JAN-1998; 98US-0072469.
XX
PA  (GENO-) GENQUEST INC.
XX
PI  Fisher PB, Zhang N;
XX
DR  WPI: 1999-458694/38.
XX  N-PSDB: AAX87383.
XX
PT  Modulation of angiogenesis by altering the expression and/or
XX  activity of a progression-associated protein, especially for cancer
XX  treatment
XX
XX  Claim 1; Page 74-76; 81pp; English.

```

```

CC The present sequence represents human progression elevated protein-3
CC (PEG-3), as deduced from a cDNA clone (see AAX87383) isolated from a
CC MCF-7 cDNA library. PEG-3 represents a novel member of the
CC gadd45/YD116 gene family. Human PEG-3 is generally expressed in
CC cells that are in progression, including most tumor cell lines. A
CC claimed method for modulating angiogenesis in an organism comprises
CC administering an agent that alters expression and/or activity of a
CC PEG-3 protein, such as an antisense polynucleotide or antibody.
CC Also claimed is a method for determining whether an agent
CC modulates angiogenesis, where the candidate agent is present
CC within a combinatorial small molecule library. Vaccines and
CC pharmaceutical compositions comprising such compounds are also
CC provided and may be used to prevent angiogenesis, especially
CC related to cancer cell progression.
XX
SQ Sequence 578 AA;

```

```

alignment_scores:
Quality: 103.00 Length: 179
Ratio: 1.198 Gaps: 11
Percent Similarity: 48.045 Percent Identity: 27.374

```

alignment_block:

US-09-651-563-808 x AA06514

Align seg 1/1 to: AA06514 from: 1 to: 578

```

112 AGTGAAGTGAAGAGAACACGACGCTTCCGAGGGTGTGTGTCAGT 161
220 ThrAspAsnLysAlaGluProSerGlySerHisSerAlaGhrPheTrpLys 236
162 GACTCAGAGTGAGAGGCCCTCGAAGTCTGTCCTTCATCATCGGTGCC 211
236 rHisThrArgGluArgProLysGlnGluGluThrLysProGluGln 253
212 ACGCCATGAGACCTTCTGTCTGTCACGGCCATACATGAGGAGGAAGA 261
253 1sArgAlaIylGlnSerHisProCysGlnAsnAlaGlnSerIuGluGly 269
262 GGGCCGAGG.....AGTGGAGGGGCTCAGGCGCAAGCTGGGGT 299
270 GlyProGluThrSerValCysSerGlySerAlaPheLeuLys..... 283
300 CTGTGGGGGTATCCGAGTCCAGAACCTGGAAACCCGACAGAGAGAT 349
284 .AlaTrpValI Tyr .Arg.....ProGlyGluAspThrGluGlu 295
350 TCTGGACTCCCA.....GACGGACACGAGAGAGGCGCATGAGGA 393
296 Glu.GluAspProAspLeuAspSerAlaGluGluAspThrAla..... 309
394 CACACACAAACACAGAACCCACACAGC..... 419
310 ..HisThrCysThrThrProHisThrSerAlaPheLeuLysAlaTrpVal 325
420 .....CAGTCCGAGAGCGCCAGTATGAGAG...CCCA 451
326 TyrArgProGlyGluAspThrGluGluAspGlyAspTrpAspSe 342
452 AAAAGAGAACACGACGCTGAAGT...CGGATCCT..... 485
342 rAlaGluGluAspAlaAlaGlnSerCysThrThrProHisThrSerAlaP 359
486 .....ACACCTGGGACGACAGACAGAAAGATGAG 515
359 heLeuLysAlaTrpValI TyrArgProGlyGluAspThrGluGluAsp 375
516 GATACAGCTGAGATCCAGTCCGCGCATGGAA 548
376 AspSerGluAsnValAlaAlaProValAspSerGlu 386

```


•
•
•
•

OM of: US-09-651-563-808 to: Pending_Patents_AA_Main: * out_format : pfs

Date: Jun 28, 2001 11:22 PM

About: Results were produced by the Gencore software, version 4.5
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```

MODEL=frame.n2p.model -DEV-XPB
-O/-c/gn12/USTPO.spool/US09651561/runat.28062001.145611.7597/app.query.fasta.1.855
-DB=Pending Patents_AA_Main -OINF=fastan -SUFFIX=ram
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPT=0.030
-LOOPEXT=0.000 -GGAPOP=4.500 -GGAPEXT=0.050 -XGAPOP=10.000
-XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -FGAPOP=10.000
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=blotsun62 -TRANS=human40.cdi -LIST=45 -DOCALLIGN=200
-THR SCORE=pcpt -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfis -NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09651561.@CN1_1_149 -NCPU=3 -LONGLOG -NO_XLPAY
-WAIT -THREADS=1

```

Search information block:

Query: US-09-651-563-808

Query Length: 781

Database sequences: 1316263

Search time (sec): 133.930000

Search time (sec): 133.930000

score_list

Score	Len	Orig	Score	Len	Documentation
/cgn2.6/ptodata/2/paa/US095.COMB	pep:US-09-560-406-809 +	887.00	1338.80	2.0e-66	1660.00
/cgn2.6/ptodata/2/paa/US095.COMB	pep:US-09-589-184-809 +	887.00	1338.80	2.0e-66	1660.00
/cgn2.6/ptodata/2/paa/US096.COMB	pep:US-09-651-563-809 +	887.00	1338.80	2.0e-66	1660.00
/cgn2.6/ptodata/2/paa/US096.COMB	pep:US-09-658-824-809 +	887.00	1338.80	2.0e-66	1660.00
/cgn2.6/ptodata/2/paa/US097.COMB	pep:US-09-671-325-809 +	887.00	1338.80	2.0e-66	1660.00
/cgn2.6/ptodata/2/paa/US097.COMB	pep:US-09-702-705-809 +	887.00	1338.80	2.0e-66	1660.00
/cgn2.6/ptodata/2/paa/US097.COMB	pep:US-09-736-457-809 +	887.00	1338.80	2.0e-66	1660.00
/cgn2.6/ptodata/2/paa/US095.COMB	pep:US-09-533-077-791 +	550.00	830.93	5.0e-38	1262.00
/cgn2.6/ptodata/2/paa/US095.COMB	pep:US-09-546-239-791 +	550.00	830.93	5.0e-38	1262.00
/cgn2.6/ptodata/2/paa/US095.COMB	pep:US-09-580-406-791 +	550.00	830.93	5.0e-38	1262.00
/cgn2.6/ptodata/2/paa/US095.COMB	pep:US-09-589-184-791 +	550.00	830.93	5.0e-38	1262.00
/cgn2.6/ptodata/2/paa/US096.COMB	pep:US-09-651-563-791 +	550.00	830.93	5.0e-38	1262.00
/cgn2.6/ptodata/2/paa/US096.COMB	pep:US-09-658-824-791 +	550.00	830.93	5.0e-38	1262.00
/cgn2.6/ptodata/2/paa/US097.COMB	pep:US-09-671-325-791 +	550.00	830.93	5.0e-38	1262.00
/cgn2.6/ptodata/2/paa/US097.COMB	pep:US-09-702-705-791 +	550.00	830.93	5.0e-38	1262.00
/cgn2.6/ptodata/2/paa/US097.COMB	pep:US-09-736-457-791 +	550.00	830.93	5.0e-38	1262.00
/cgn2.6/ptodata/2/paa/US06.COMB	pep:US-60-147-409-5605 +	536.00	805.81	8.2e-37	99.97
/cgn2.6/ptodata/2/paa/US097.COMB	pep:US-09-702-705-1831 +	507.00	766.38	2.0e-34	1089.96
/cgn2.6/ptodata/2/paa/US097.COMB	pep:US-09-736-457-1831 +	507.00	766.38	2.0e-34	1089.96
/cgn2.6/ptodata/2/paa/US095.COMB	pep:US-09-518-642-786 +	423.50	640.89	2.2e-27	1089.96
/cgn2.6/ptodata/2/paa/US095.COMB	pep:US-09-533-077-786 +	423.50	640.89	2.2e-27	1089.96
/cgn2.6/ptodata/2/paa/US095.COMB	pep:US-09-543-077-793 +	423.50	640.89	2.2e-27	1089.96
/cgn2.6/ptodata/2/paa/US095.COMB	pep:US-09-556-279-786 +	423.50	640.89	2.2e-27	1089.96
/cgn2.6/ptodata/2/paa/US095.COMB	pep:US-09-546-239-793 +	423.50	640.89	2.2e-27	1089.96
/cgn2.6/ptodata/2/paa/US095.COMB	pep:US-09-560-406-786 +	423.50	640.89	2.2e-27	1089.96
/cgn2.6/ptodata/2/paa/US095.COMB	pep:US-09-560-406-793 +	423.50	640.89	2.2e-27	1089.96
/cgn2.6/ptodata/2/paa/US095.COMB	pep:US-09-589-184-786 +	423.50	640.89	2.2e-27	1089.96
/cgn2.6/ptodata/2/paa/US095.COMB	pep:US-09-589-184-793 +	423.50	640.89	2.2e-27	1089.96
/cgn2.6/ptodata/2/paa/US096.COMB	pep:US-09-651-563-786 +	423.50	640.89	2.2e-27	1089.96
/cgn2.6/ptodata/2/paa/US096.COMB	pep:US-09-651-563-793 +	423.50	640.89	2.2e-27	1089.96
/cgn2.6/ptodata/2/paa/US096.COMB	pep:US-09-658-824-786 +	423.50	640.89	2.2e-27	1089.96
/cgn2.6/ptodata/2/paa/US096.COMB	pep:US-09-658-824-793 +	423.50	640.89	2.2e-27	1089.96
/cgn2.6/ptodata/2/paa/US097.COMB	pep:US-09-671-325-786 +	423.50	640.89	2.2e-27	1089.96
/cgn2.6/ptodata/2/paa/US097.COMB	pep:US-09-671-325-793 +	423.50	640.89	2.2e-27	1089.96
/cgn2.6/ptodata/2/paa/US097.COMB	pep:US-09-702-705-786 +	423.50	640.89	2.2e-27	1089.96
/cgn2.6/ptodata/2/paa/US097.COMB	pep:US-09-702-705-793 +	423.50	640.89	2.2e-27	1089.96
/cgn2.6/ptodata/2/paa/US097.COMB	pep:US-09-736-457-786 +	423.50	640.89	2.2e-27	1089.96
/cgn2.6/ptodata/2/paa/US097.COMB	pep:US-09-736-457-793 +	423.50	640.89	2.2e-27	1089.96
/cgn2.6/ptodata/2/paa/US097.COMB	pep:US-09-762-705-1830 +	380.00	579.93	9.4e-24	66.66
/cgn2.6/ptodata/2/paa/US097.COMB	pep:US-09-736-457-1830 +	380.00	579.93	9.4e-24	66.66

/cgn2_6/protdata/2/paa/USO95_COMB.pep	:US-09-560-406-809
/cgn2_6/protdata/2/paa/USO95_COMB.pep	:US-60-197-873-19905 + 251.00 381.91 7.5e-13
/cgn2_6/protdata/2/paa/USO94_COMB.pep	:US-09-488-725A.2733 + 251.00 379.53 7.9e-13
/cgn2_6/protdata/2/paa/USO6_COMB.pep	:US-60-117-873-25416 + 251.00 379.53 7.9e-13
/cgn2_6/protdata/2/paa/USO6_COMB.pep	:US-60-238-331-127 + 251.00 379.53 7.9e-13
/cgn2_6/protdata/2/paa/USO6_COMB.pep	:US-60-277-380-36 + 251.00 379.53 7.9e-13

```
seq.documentation_block:
: Sequence 809, Application US/09560406
: GENERAL INFORMATION:
: APPLICANT: Wang, Tongtong
: APPLICANT: Bangur, Chaitanya S.
: APPLICANT: Lodes, Michael A.
: APPLICANT: Fanger, Gary
: APPLICANT: Veddyick, Tom
: APPLICANT: Carter, Darlick
: APPLICANT: Retter, Marc
: APPLICANT: Mannheim, Jane
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
: TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
: FILE REFERENCE: 210121.478c7
: CURRENT APPLICATION NUMBER: US/09/560,406
: NUMBER OF SEQ ID NOS: 824
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 809
: LENGTH: 160
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-560-406-809

alignment_scores:
    quality: 887.00      Length: 160
    Ratio: 5.54         Gaps: 0
    Percent Similarity: 100.000   Percent Identity: 100.000

Alignment Block:
US-09-651-563-808 x US-09-560-406-809 ..

Align seg 1/1 to: US-09-560-406-809 from: 1 to: 160

202 ATGCGGTGGCCAGGCCCATGGACTTCTGTCTGCACAGGCCATTAAGTAG 251
|||||
1 MetAdgCyuhIsAlaHIsGLyProSerCyusLeuValThrAlaIlePheArg 17

252 GGAGGAGGAGGGGCCGAGAGTAGTGAGGGGGCTCAAGGCCAAGCTGGGTGCT 301
|||||
17 gctugtcgtgcgtatccggatccgaagacacttgaaaccgccgacagaagtcc 351
|||||
302 GTTTGGGGGTATCCGAGTCCGAAACACTGTAAGCCCGGACAGAAGATTTC 351
|||||
34 ysttptdlytyrproserProMrgSerPrnTrpAspProAspArgPhe 50
|||||
352 TSGACTCCCCAGACGCGACGACGAGAGGAGGAGCGCATGACGACACACACA 401
|||||
51 tprtlrrproglnthrGLyProGLyLgnlgrnhlsGLyadgrnhlsIngl 67

402 AACGACAGAACCCACACAGCCAGTCCCAGAGGCCAGTAATGAGAGCCCCA 451
|||||
67 nhrhglnsnhnslhrAlaSerProArgSerProValMeteluberProL 84
|||||
452 AAAAGAGAACGACAGAGCTGAAAGTCGGGATCTTACACCTGGACGACAGA 501
|||||
84 yblsylysAnslnlnleuylsValIGlylleuHnsleuGLySerArg 1000
|||||
502 CAGAAGAACATCAGGATACAGCTGAGATCCACGTGCGGACATGGAAGT 551
|||||
101 GlnLylysllleargllleglnleuArgserGncysalaThrTrpLysVa 117
|||||
552 GATTCGCAAGAGCTGATCATGTCMAACACCGGGGATTAATGTGATTGG 601
|||||
```

```
117 lilecylsersercysllesercInthrProgllylleasnleuAspleug 134
602 GTTCCGGCGCTCAAGGTGAAGTAATACCTAAAGAGAAACACTGTAAATG 651
134 lYsercilyallysvallyslleleProlysglulhiscylslysmet 150
652 CCAGAAAGCAGGTGAAGAGCAACCAAGTT 681
151 ProclualaglygluluglnProglInval 160

seq.name: /cgn2_6/ptodata/2/paa/US095_COMB.pep:US-09-589-184-809
seq.documentation_block:
: Sequence 809, Application US/09589184
: GENERAL INFORMATION:
: APPLICANT: Wang, Tongtong
: APPLICANT: Bangur, Chaitanya S.
: APPLICANT: Lodes, Michael A.
: APPLICANT: Fanger, Gary
: APPLICANT: Vedvick, Tom
: APPLICANT: Carter, Darriick
: APPLICANT: Retter, Marc
: APPLICANT: Mannion, Jane
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
: FILE REFERENCE: 210121.478C8
: CURRENT APPLICATION NUMBER: US/09/589,184
: CURRENT FILING DATE: 2000-06-05
: NUMBER OF SEQ ID NOS: 827
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 809
: LENGTH: 160
: TYPE: PRF
: ORGANISM: Homo sapiens
US-09-589-184-809

alignment_scores:
Quality: 887.00 Length: 160
Ratio: 5.544 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-651-563-808 x US-09-589-184-809
Align seg 1/1 to: US-09-589-184-809 from: 1 to: 160
202 ATGCGGTGCGACGCCATGAGCTTCTGTCTGCTCAGCGCATACCTAG 251
1 MetatrgcYshIsAlahIsclYProserCysleuValThrAlaIleThrAr 17
252 GGAGGAAGAGGGGCCGAGAGTGAAGGGCTCAGCGGAAGCTGGGGTCT 301
17 gclugluluglylProArYserclYclYAlaIalAlalysleuglyCysC 34
302 GTTGGGGGTATCCGAGTCCAGAACCACTGGAACCCCGACAGAAGATTG 351
34 ystrpGlYtrProserProArYserThrTrpAsnProAspArgpHe 50
352 TGGACTCCCCAGACGGGACGAGAGAGGAGCGCATGACGACACACACA 401
51 TrpThrProglInthrGlYProglYglulYArghIsgluArghIsThrGl 67
402 AACACAGAACCACACAGCCAGTCCAGAGGCCAGTATGAGAGCCCA 451
67 nThrGlnAsnHsThrAlaSerProArYserProValMetGluserProL 84
452 AAAAGAAGAACCCAGAGCTGAAGTCGGAGATCTACACTGGGACGACAGA 501
84 yslYslYAsnclnGlnleuYslValglYlIleleuHIsleuglySerArg 100
502 CAGAAGAAGATCAGATACAGCTGAGATCCCAAGTCGGGACATGGAAGCT 551
```

```
101 GlnYslYslleArqllleGlnleuArYserClnCysAlaThrTrpYsVa 117
552 GATTCGACAGAGCTGATCATGATAACACCCGGGAGTAATCTGATTGG 601
117 lilecylsersercysllesercInthrProgllylleasnleuAspleug 134
602 GTTCCGGCGCTCAAGGTGAAGTAATACCTAAAGAGAAACACTGTAAATG 651
134 lYsercilyallysvallyslleleProlysglulhiscylslysmet 150
652 CCAGAAAGCAGGTGAAGAGCAACCAAGTT 681
151 ProclualaglygluluglnProglInval 160

seq.name: /cgn2_6/ptodata/2/paa/US096_COMB.pep:US-09-651-563-809
seq.documentation_block:
: Sequence 809, Application US/09651563
: GENERAL INFORMATION:
: APPLICANT: Wang, Tongtong
: APPLICANT: Bangur, Chaitanya S.
: APPLICANT: Lodes, Michael A.
: APPLICANT: Fanger, Gary
: APPLICANT: Vedvick, Tom
: APPLICANT: Carter, Darriick
: APPLICANT: Retter, Marc
: APPLICANT: Mannion, Jane
: APPLICANT: Fan, Liqun
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
: FILE REFERENCE: 210121.478C10
: CURRENT APPLICATION NUMBER: US/09/651,563
: CURRENT FILING DATE: 2000-08-29
: NUMBER OF SEQ ID NOS: 1679
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 809
: LENGTH: 160
: TYPE: PRF
: ORGANISM: Homo sapiens
US-09-651-563-809

alignment_scores:
Quality: 887.00 Length: 160
Ratio: 5.544 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-651-563-808 x US-09-651-563-809
Align seg 1/1 to: US-09-651-563-809 from: 1 to: 160
202 ATGCGGTGCGACGCCATGAGCTTCTGTCTGCTCAGCGCATACCTAG 251
1 MetatrgcYshIsAlahIsclYProserCysleuValThrAlaIleThrAr 17
252 GGAGGAAGAGGGGCCGAGAGTGAAGGGCTCAGCGGAAGCTGGGGTCT 301
17 gclugluluglylProArYserclYclYAlaIalAlalysleuglyCysC 34
302 GTTGGGGGTATCCGAGTCCAGAACCACTGGAACCCCGACAGAAGATTG 351
34 ystrpGlYtrProserProArYserThrTrpAsnProAspArgpHe 50
352 TGGACTCCCCAGACGGGACGAGAGGAGCGCATGACGACACACACA 401
51 TrpThrProglInthrGlYProglYglulYArghIsgluArghIsThrGl 67
402 AACACAGAACCACACAGCCAGTCCAGAGGCCAGTATGAGAGCCCA 451
67 nThrGlnAsnHsThrAlaSerProArYserProValMetGluserProL 84
452 AAAAGAAGAACCCAGAGCTGAAGTCGGAGATCTACACTGGGACGACAGA 501
```



```

202 ATGGCGTCCAGCCGACCTGCTTCTGTCACGCGCATTAACCTAG 251
1 MetArgYshLSAlhSLgIyProSerCyLeuValThrAlaIleThrAr 17
252 GGAGGAAGGAGGCGCCAGAGTGGAGGGGCTCAGCGCAACCTGGGTCT 301
17 gGIuGIuGIyGIyProArSerGIyGIyAlaGIuAlaLySLeugIyCySc 34
302 GTTGGGGGTTCGAGTCCGAGAACGACCTGGAACCCCGACAGAAATTC 351
34 ySTpGIyTyProSerProArSerThrTrpAsnProAspAlaArgPhe 50
352 TGGACTCCCGAGAGCGGACGAGAGGAGGAGCGCATGAGGACACACA 401
51 TrpThrProGIuInThrGIyProGIyGIuGIyArgHisGIuArgHisThrGI 67
402 AACGACAAACCCACAGCCAGCTCCAGAGCCCGCATTAATGGAGGCCCA 451
67 nThrGIuAsnHisThrAlaSerProArSerProValMetGIuSerProL 84
452 AAAGAGAACCCAGCAGCTGAAAGTCGGATCTTACACTGGCGACAGACA 501
84 ySLySLySAsnGIuInLeuLySLyAlGIyIleLeuHisLeuGIySerArg 100
502 CAGAGAAGATCAGGATPACAGCTGAGATCCCACTGCCGACATGGAAGT 551
101 GIuLySLySLyIleArgIleGIuInLeuArgSerGIuInCysAlaThrTrpLySva 117
552 GATCTGCAAGAGCTGCATCAAGCTCAACACCGGGATTAATCTGGATTGG 601
117 ILeCysLySerCysIleSerGIuInThrProGIyIleAsnLeuAspLeuG 134
602 GTTCCGGCGTCAGAGTGAAGTAAATTAATTAAGAGAACACTGTAATAG 651
134 LySerGIyValLySLyValLySLyIleIleProLySLyGIuInHisCysLyMet 150
652 CCAGAGCAGGTGAAGAGCAACCAAGTT 681
151 ProGIuInAGIyGIuInGIuInProGIuInVal 160

```

seq_name: /cgn2_6/prodata/2/paa/US095_COMB.pep:US-09-533-077-791

```

seq_documentation_block:
; Sequence 791, Application US/09533077
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C5
; CURRENT APPLICATION NUMBER: US/09/533,077
; CURRENT FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 800
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 791
; LENGTH: 126
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-533-077-791

```

alignment_scores:

```

Quality: 550.00 Length: 227
Ratio: 4.583 Gaps: 2
Percent Similarity: 52.863 Percent Identity: 52.423

```

alignment_block:

```

US-09-651-563-808 x US-09-533-077-791 ..
Align seg 1/1 to: US-09-533-077-791 from: 1 to: 126

```

```

2 CGGCGAGCTGTAGGCCGCGACCTGGGCTCCCTGAGTCTGATCTTTC 51
7 ArgArSerCySLuProAlaThrArgValProGIuValThrIleLeuSe 23
52 TCCGCTACTGAGAACACGGGGGTAGGTCCACAGGAGATCCAACTGGGAG 101
23 rProLeuLeuArgHisGLy..... 29
102 TTGAAGTGTGAGTGAAGACTGAAGAGAACACAGGAGCTTCCGAGGGTTG 151
29 ..... 29
152 TGTGTCACTAGCTAGACTGAGAAAGCCCTGAAAGTGTCTCCCTCTC 201
29 ..... 29
202 ATGCGGTGCGACGCCCATGAGACCTTCTGTCTGTACAGGCGCATAGTAG 251
29 ..... 29
252 GGAGGAAGGAGGCGCGAGAGTGTGAGGGGCTCAGGCGAAGCTGGGGTGT 301
29 ..... 29
302 GTTGGGGGTATCCGAGTCCGACAGACCTGGAACCCCGACAGAAATTC 351
29 ..... 29
352 TGGACTCCCGAGAGCGGACGAGAGGAGGAGCGCATGAGCGACACACA 401
30 ..... Gly..HisThrGI 33
402 AACGACAAACCCACAGCCAGCTCCAGAGCCCGCATTAATGGAGGCCCA 451
33 nThrGIuAsnHisThrAlaSerProArSerProValMetGIuSerProL 50
452 AAAGAGAACCCAGCAGCTGAAAGTCGGATCTTACACTGGCGACAGACA 501
50 ySLySLySAsnGIuInLeuLySLyAlGIyIleLeuHisLeuGIySerArg 66
502 CAGAGAAGATCAGGATPACAGCTGAGATCCCACTGCCGACATGGAAGT 551
67 GIuLySLySLyIleArgIleGIuInLeuArgSerGIuInCysAlaThrTrpLySva 83
552 GATCTGCAAGAGCTGCATCAAGCTCAACACCGGGATTAATCTGGATTGG 601
83 ILeCysLySerCysIleSerGIuInThrProGIyIleAsnLeuAspLeuG 100
602 GTTCCGGCGTCAGAGTGAAGTAAATTAATTAAGAGAACACTGTAATAG 651
100 LySerGIyValLySLyValLySLyIleIleProLySLyGIuInHisCysLyMet 116
652 CCAGAGCAGGTGAAGAGCAACCAAGTT 681
117 ProGIuInAGIyGIuInGIuInProGIuInVal 126

```

seq_name: /cgn2_6/prodata/2/paa/US095_COMB.pep:US-09-546-259-791

```

seq_documentation_block:
; Sequence 791, Application US/09546259
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C6
; CURRENT APPLICATION NUMBER: US/09/546,259
; CURRENT FILING DATE: 2000-04-10
; NUMBER OF SEQ ID NOS: 803
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 791

```

; LENGTH: 126
; TYPE: PR
; ORGANISM: Homo sapiens
US-09-546-259-791

alignment_scores:

Quality: 550.00 Length: 227
Ratio: 4.583 Gaps: 2
Percent Similarity: 52.863 Percent Identity: 52.423

alignment_block:

US-09-651-563-808 x US-09-546-259-791 ..

Align seg 1/1 to: US-09-546-259-791 from: 1 to: 126

```
2 CGCGGAGCTGTAGCCGCGGCTCCCTGAGTCTGATTTCTTC 51
|||||
7 ArgArgSerCysGluProAlaThrArgValProGluValTrrIleLeu 23
|||||
52 TCCGCTACTGAGACAGCGGGGTAGTCCACAGGACGATCCACATGGGAG 101
|||||
23 rProLeuLeuArgHisGly..... 29
102 TTGAAGTGTAGTGAAGTGAAGAACACAGCAGCTTCGGAGGTTG 151
..... 29
152 TTGTGCTAGTCACTAGATGAGAACGCCCTCGAAGTGTGCTCCCTTC 201
..... 29
202 ATGCGGTGACACGCCCATGACCTTCTTGTCTGTCGACGCGCATTAAGTAG 251
..... 29
252 GGAGAGAGAGGGCCGAGAGTGAAGGGCTCAGCGAAGCTGGGGTCT 301
..... 29
302 GTTGGGGGTATCCGAGTCCAGAACACCTGGAACCCCGACAGAAAGATT 351
..... 29
29 ..... 29
352 TGGACTCCCAACGCGGACCAAGAGGGAGCGCATGACGACACACACA 401
|||||
30 .....Gly..HisThrG1 33
402 AACACAGAACACACAGCCAGTCCAGAGAGCCCATTAATGAGAGCCCA 451
|||||
33 nThrGlnAsnHisThrAlaSerProArgSerProValMetGluSerProL 50
452 AAAAGAAAGACAGCAGCTGAAGTGGGATCTTACACCTGGGAGAGAGA 501
|||||
50 ySlySlyAsnGlnGlnLeuLysValGlyIleLeuHisLeuGlySerArg 66
66 GlnySlySlyIleArgIleGlnLeuArgSerGlnCysAlaThrTrpLysVa 83
83 GATCGCAGAGCTGCATCACTCAACACCGGGGATTAATCTGATTTGG 601
|||||
552 GATCGCAGAGCTGCATCACTCAACACCGGGGATTAATCTGATTTGG 601
|||||
83 lIleCysSlySerCysIleSerGlnThrProGlyIleAsnLeuAspLeuG 100
602 GTTCGCGGCTCAAGGTGAAGATTAATACCTTAAGAGAACTGTAAATG 651
|||||
100 ySerGlyValLysValLysIleLeuProLysGluGluHisCysLysMet 116
652 CCAGAGCAGGTGAAGAGACCAACCAAGTT 681
|||||
117 ProGluAlaGlyGluGluGlnProGluVal 126
```

seq_name: /cgn2_6/prodata/2/paa/us095_COMB.pep.us-09-560-406-791

seq_documentation_block:

; Sequence 791, Application US/09560406

; GENERAL INFORMATION:

; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darlick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE REFERENCE: 210121.478C7
; CURRENT APPLICATION NUMBER: US/09/560,406
; NUMBER OF SEQ ID NOS: 824
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 791
; LENGTH: 126
; TYPE: PR
; ORGANISM: Homo sapiens
US-09-560-406-791

alignment_scores:

Quality: 550.00 Length: 227
Ratio: 4.583 Gaps: 2
Percent Similarity: 52.863 Percent Identity: 52.423

alignment_block:

US-09-651-563-808 x US-09-560-406-791 ..

Align seg 1/1 to: US-09-560-406-791 from: 1 to: 126

```
2 CGCGGAGCTGTAGCCGCGGCTCCCTGAGTCTGATTTCTTC 51
|||||
7 ArgArgSerCysGluProAlaThrArgValProGluValTrrIleLeu 23
|||||
52 TCCGCTACTGAGACAGCGGGGTAGTCCACAGGACGATCCACATGGGAG 101
|||||
23 rProLeuLeuArgHisGly..... 29
102 TTGAAGTGTAGTGAAGTGAAGAACACAGCAGCTTCGGAGGTTG 151
..... 29
152 TTGTGCTAGTCACTAGATGAGAACGCCCTCGAAGTGTGCTCCCTTC 201
..... 29
202 ATGCGGTGACACGCCCATGACCTTCTTGTCTGTCGACGCGCATTAAGTAG 251
..... 29
29 ..... 29
252 GGAGAGAGAGGGCCGAGAGTGAAGGGGCTCAGGCGAAGTGGGGTCT 301
..... 29
29 ..... 29
302 GTTGGGGGTATCCGAGTCCAGAACACCTGGAACCCCGACAGAAAGATT 351
..... 29
29 ..... 29
352 TGGACTCCCAACGCGGACCAAGAGAGGAGCGCATGACGACACACACA 401
|||||
30 .....Gly..HisThrG1 33
402 AACACAGAACACACAGCCAGTCCAGAGAGCCCATTAATGAGAGCCCA 451
|||||
33 nThrGlnAsnHisThrAlaSerProArgSerProValMetGluSerProL 50
452 AAAAGAAAGACAGCAGCTGAAGTGGGATCTTACACCTGGGAGAGAGA 501
```



```

|||||
50 ysllyslsasnnglnglnleulysvalglylleuuhlsleuglyserarg 66
502 CAGAGAGAGTGCAGATACAGCTGAGATCCCGAGTGGCGGACATGAGAGT 551
67 glnlyslsylearlgileglnleuaryserglnlcysalathttrplysya 83
552 GATCTGCAGAGCTGCATCAGTCAACACCGGGGATTAATCTGATTTGG 601
83 lilecylsersercysilleserglnthrproglylleasnlleuaspheug 100
602 GTTCCGGCGTCAGGTGAAGATTAATACCTAAAGAGAACACTGTAATG 651
100 lyserglyvallyslsvallyslleleprolysglunhiscyslysmet 116
652 CCAGAGCAGGTGAAGAGCAACCAAGTT 681
117 Proglunlaaglyglunhiscyslysmet 126

```

seq_name: /cgn2_6/ptodata/2/paa/US095_COMB.pep:US-09-589-184-791

seq_documentation_block:

```

; Sequence 791, Application US/09589184
; GENERAL INFORMATION:
; APPLICANT: Wang, Tonglong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodges, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE REFERENCE: 210121.478C8
; CURRENT APPLICATION NUMBER: US/09/589,184
; NUMBER OF SEQ ID NOS: 827
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 791
; LENGTH: 126
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-589-184-791

```

alignment_scores:

Quality:	550.00	Length:	227
Ratio:	4.583	Gaps:	2
Percent Similarity:	52.863	Percent Identity:	52.423

alignment_block:

US-09-651-563-808 x US-09-589-184-791 ..

Align seg 1/1 to: US-09-589-184-791 from: 1 to: 126

```

2 CGGGGAGAGTGTGAGCCGGGAGCTCGGGCTCGAGGCTGTGATTTCTTC 51
7 Argtrsercysglunprolatrharvalproglunvaltrpleuse 23
52 TCCGCTACTAGACAGCGCGGGTACGTCACAGGACAGATCCAACCTGGAG 101
23 rProleuleuarghisgly..... 29
102 TTGAAGTGTGAGTGAAGTAAGAGAACACAGAGCTCCGAGGGTTG 151
29 ..... 29
152 TGTGTCAGTACTCAGAGTGAAGAGCCCTCGAAGTGTGTCCCTCTC 201
29 ..... 29
202 ATGGGTCACAGCCCATGAGACTTCTTGTCTGTACGGCATAACTAG 251

```

```

29 ..... 29
252 GGAGAGAGAGGGCCGAGAGATGAGAGGGCTCAGGCGAAGTGGGGTCT 301
29 ..... 29
302 GTTGGGGTATCCAGATCCAGAGACACCTGTGMAACCCGACAGAGATTG 351
29 ..... 29
352 TGGACTCCCGACAGGGGACCCAGAGAGGAGCGCATGAGCCGACAGACA 401
30 ..... 401
402 AACACAGACACACAGCCAGTCCAGAGAGCCAGTAATGAGAGCCCA 451
33 nhrghlnasnhtthralsrprohrgserprovalmetgluserprol 50
452 AAAGAAGAACAGCAGCTGAAGATCGGGATCTACACCTGGGAGACAGA 501
50 ysllyslsasnnglnglnleulysvalglylleuuhlsleuglyserarg 66
502 CAGAGAGATCAGGATACAGTCCAGATCCCGAGTGGCGGACATGGAAGT 551
67 glnlyslsylearlgileglnleuaryserglnlcysalathttrplysya 83
552 GATCTGCAGAGCTGCATCAGTCAACACCGGGGATTAATCTGATTTGG 601
83 lilecylsersercysilleserglnthrproglylleasnlleuaspheug 100
602 GTTCCGGCGTCAGGTGAAGATTAATACCTAAAGAGAACACTGTAATG 651
100 lyserglyvallyslsvallyslleleprolysglunhiscyslysmet 116
652 CCAGAGCAGGTGAAGAGCAACCAAGTT 681
117 Proglunlaaglyglunhiscyslysmet 126

```

seq_name: /cgn2_6/ptodata/2/paa/US096_COMB.pep:US-09-651-563-791

seq_documentation_block:

```

; Sequence 791, Application US/09651563
; GENERAL INFORMATION:
; APPLICANT: Wang, Tonglong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodges, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE REFERENCE: 210121.478C10
; CURRENT APPLICATION NUMBER: US/09/651,563
; NUMBER OF SEQ ID NOS: 1679
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 791
; LENGTH: 126
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-651-563-791

```

alignment_scores:

Quality:	550.00	Length:	227
Ratio:	4.583	Gaps:	2
Percent Similarity:	52.863	Percent Identity:	52.423

alignment_block:

US-09-651-563-808 x US-09-651-563-791 ..

Align seg 1/1 to: US-09-651-563-791 from: 1 to: 126

```

2 CGCGGAGCTGTGAGCCGGGAGCTGGGCTCTGAGTGTGATTTTC 51
7 ArgArgSerCysGlnProAlaThrArgValProGluValTrrPleuSe 23
52 TCCGCTACTGAGACACGCGGGTAGGTCCACAGCAGATCCAACTGGGAG 101
23 rProleuLeuArgHisGly..... 29
102 TTGAAGTGTGAGTGAAGAGGAGAACACAGGCTTCGGAGGGTTG 151
29 ..... 29
152 TGTGTCACTGACTCAGAGTAGAAGGCCCTGAAAGTCGTGCTCCTTC 201
29 ..... 29
202 ATGCGGTGCCAGCCCATGACCTTCTTCTGTCAGGCGCATACTAG 251
29 ..... 29
252 GGAGAGAGGAGGCGGAGAGTGAAGGGCTCAGCGGAAGTGGGGTCT 301
29 ..... 29
302 GTTGGGGGTATCCGAGTCCCAAGACACTGGAACCCGACAGAAATTC 351
29 ..... 29
352 TGAGTCCCCAGACGGGACCCAGAGAGGAGCGCATGAGCGACACACA 401
30 ..... 33
402 AACACAGAACACACAGCCAGTCCAGAGAGCCAGTAATGAGAGCCCA 451
33 nThrGlnAsnHisThrAlaSerProArgSerProValMetGlnSerPro 50
452 AAAAGAGAACGACGATGAAAGTCGGGATCTTACCTGGGCGACGAGA 501
50 yslYslYsAsnGlnGlnLeuYsValGlyTlleuHisLeuGlySerArg 66
552 GATCTGCAAGAGCTGCATCAGTCAAAACACCGGGATTAATCTGATTGG 601
67 GlnYslYsIleArgTlleGlnLeuArgSerGlnCysAlaThrTrpLysVa 83
83 lIleCysLysSerCysIleSerGlnThrProGlyIleAsnLeuAspLeuG 100
602 GTTCCGGCTCAAGGTGAAGATATACCTAAAGAGAGCACTGTAATG 651
100 YserGlyValIlyValIlyIleIleProLysGlnGlnHisCysLysMet 116
652 CCAGAGAGCAGGTGAAGAGCAACCAAGTT 681
117 ProGlnAlaGlyGlnGlnGlnProGlnVal 126
seq_name: /cgn2_6/ptodata/2/paa/US096_COMB.pep:US-09-658-824-791
seq_documentation_block:
: Sequence 791, Application US/09658824
: GENERAL INFORMATION:
: APPLICANT: Bangur, Tongtong
: APPLICANT: Lodes, Michael A.
: APPLICANT: Fanger, Gary
: APPLICANT: Vedvik, Tom
: APPLICANT: Carter, Darick
: APPLICANT: Relfer, Marc
: APPLICANT: Mannion, Jane

```

```

: APPLICANT: Fan, Liqun
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
: TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
: FILE REFERENCE: 210121.478C11
: CURRENT APPLICATION NUMBER: US/09/658,824
: NUMBER OF SEQ ID NOS: 1788
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 791
: LENGTH: 126
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-658-824-791

```

alignment_scores: Quality: 550.00 Length: 227
Ratio: 4.583 Gaps: 2
Percent Similarity: 52.863 Percent Identity: 52.423

alignment_block:

US-09-651-563-808 x US-09-658-824-791 ..

Align seg 1/1 to: US-09-658-824-791 from: 1 to: 126

```

2 CGCGGAGCTGTGAGCCGGGAGCTGGGCTCTGAGTGTGATTTTC 51
7 ArgArgSerCysGlnProAlaThrArgValProGluValTrrPleuSe 23
52 TCCGCTACTGAGACACGCGGGTAGGTCCACAGCAGATCCAACTGGGAG 101
23 rProleuLeuArgHisGly..... 29
102 TTGAAGTGTGAGTGAAGAGGAGAACACAGGCTTCGGAGGGTTG 151
29 ..... 29
152 TGTGTCACTGACTCAGAGTAGAAGGCCCTGAAAGTCGTGCTCCTTC 201
29 ..... 29
202 ATGCGGTGCCAGCCCATGACCTTCTTCTGTCAGGCGCATACTAG 251
29 ..... 29
252 GGAGAGAGGAGGCGGAGAGTGAAGGGCTCAGGCGGAAGCTGGGGTCT 301
29 ..... 29
302 GTTGGGGGTATCCGAGTCCAGAAACACTGGAACCCGACAGAAATTC 351
29 ..... 29
352 TGAGTCCCCAGACGGGACCCAGAGAGGAGCGCATGAGCGACACACA 401
30 ..... 33
402 AACACAGAACACACAGCCAGTCCAGAGAGCCAGTAATGAGAGCCCA 451
33 nThrGlnAsnHisThrAlaSerProArgSerProValMetGlnSerPro 50
452 AAAAGAGAACGACGATGAAAGTCGGGATCTTACCTGGGCGACGAGA 501
50 yslYslYsAsnGlnGlnLeuYsValGlyTlleuHisLeuGlySerArg 66
502 CAGAGAGAGTCAAGATACAGCTGAGATCCCAATGGGCGACATGAGAGT 551
67 GlnYslYsIleArgTlleGlnLeuArgSerGlnCysAlaThrTrpLysVa 83
552 GATCTGCAAGAGCTGCATCAGTCAAAACACCGGGATTAATCTGATTGG 601
83 lIleCysLysSerCysIleSerGlnThrProGlyIleAsnLeuAspLeuG 100

```

```

602 GTTCCGGCTCAAGTGAAGATATACCTAAAGAGAGACACTGTAAATG 651
100 lyserglyallysvalyslleleprolysgluinhiscyslysmet 116
652 CCAGAGCAGGTGAAGAGCACCACCAAGTT 681
117 ProgluAlaGlygluInProgluInval 126

```

seq_name: /cgn2_6/ptodata/2/paa/US096_COMB.pep:US-09-671-325-791

seq_documentation_block:

```

; Sequence 791, Application US/09671325
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darlick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C12
; CURRENT APPLICATION NUMBER: US/09/671.325
; NUMBER OF SEQ ID NOS: 1825
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 791
; LENGTH: 126
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-671-325-791

```

alignment_scores:

```

Quality: 550.00 Length: 227
Ratio: 4.583 Gaps: 2
Percent Similarity: 52.863 Percent Identity: 52.423

```

alignment_block:

US-09-651-563-808 x US-09-671-325-791 ..

Align seg 1/1 to: US-09-671-325-791 from: 1 to: 126

```

2 CGGCGAGCTGTGAGCGCGGAGCTCGGTCCTCGAGCTGTGATTTTC 51
7 ArgArgserCysgluProAlaThrArgValProgluValTrpIleLeuSe 23
52 TTCGCTACTGAGACAGCGCGGTAGGTCCACAGGAGATCCAACTGGAG 101
23 rProLeuLeuArgHisGly..... 29
102 TTGAAGTGTGAGTGAAGTGAAGAGAACCAAGCAGGCTTCGGAGGGTTG 151
29 ..... 29
152 TGGGTGACTGACTCAGAGTGAAGAGCCCTCGAATCGTCGCTCTTC 201
29 ..... 29
202 ATGCGGTGACGCGCCATGACCTTCTGTCTGTCACGCGCATTAAG 251
29 ..... 29
252 GAGAGAGAGGCGCCAGAGTGAAGGGCTCAGCGAACTGGGGTGTCT 301
29 ..... 29
302 GTTGGGGGTATCGAGTCCAGAGACCTGGAAACCCGAGAAAGATTTC 351
29 ..... 29

```

```

352 TGGACTCCCAAGCGGAGCCAGAGAGGGGAGCATGAGCCACACACA 401
30 .....Gly..HisTrpI 33
402 AACACAGAACCCACAGCAGCTCCAGGAGCCAGTAATGAGAGCCCA 451
33 ThrGlnAsnHisThrAlaSerProAlaSerProValMetIleSerProI 50
452 AAAGAGAACCCAGCAGCTGAAAGTCGGGATCTTACACCTGGCAGAGA 501
50 yslYslYlsAsnGlnInleuYslValGlyIleLeuHisLeuGlySerArg 66
502 CAGAGAAAGATCAGATACAGTGAATCCAGTCCAGTCCGAGATGAAAG 551
67 GlnYslYslLeArgIleGlnleuArgSerGlnCysAlaThrTrpYlsVa 83
552 GATTCGCAAGCTGCATCAGTCAACACCGGGGATTAATGTGATTGG 601
83 lIleCysYlsSerCysIleSerGlnThrProGlyIleAsnleuAspLeuG 100
602 GTTCCGGCTCAAGTGAAGATATACCTAAAGAGAGACACTGTAAATG 651
100 lyserglyallysvalyslleleprolysgluinhiscyslysmet 116
652 CCAGAGCAGGTGAAGAGCACCACCAAGTT 681
117 ProgluAlaGlygluInProgluInval 126

```

seq_name: /cgn2_6/ptodata/2/paa/US097_COMB.pep:US-09-702-705-791

seq_documentation_block:

```

; Sequence 791, Application US/09702705
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darlick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C14
; CURRENT APPLICATION NUMBER: US/09/702.705
; NUMBER OF SEQ ID NOS: 1833
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 791
; LENGTH: 126
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-702-705-791

```

alignment_scores:

```

Quality: 550.00 Length: 227
Ratio: 4.583 Gaps: 2
Percent Similarity: 52.863 Percent Identity: 52.423

```

alignment_block:

US-09-651-563-808 x US-09-702-705-791 ..

Align seg 1/1 to: US-09-702-705-791 from: 1 to: 126

```

2 CGGCGAGCTGTGAGCGCGGAGCTCGGTCCTCGAGTGTGATTTTC 51
7 ArgArgserCysgluProAlaThrArgValProgluValTrpIleLeuSe 23
52 TTCGCTACTGAGACAGCGCGGTAGGTCCACAGGAGATCCAACTGGAG 101

```

```
23 rProLeuLeuArgHisGly..... 29
102 TTGAAGTGTGAGTGAAGTGAAGAGAACACAGAGCTTCGGAGGGTTG 151
29 ..... 29
152 TGTGGTCAGTGACTGAGTGAAGAGCCCTGGAAGTGTGTCCTCTC 201
29 ..... 29
202 ATGCGGTGCCAGCCCATGACCTTCTGTCTGTCTACAGGCCATTAAGTAG 251
29 ..... 29
252 GGAGGAGAGAGGGCCGAGAGAGTGAAGGGGCTCAGGCGAAGCTGGGGTCT 301
29 ..... 29
302 GTTGGGGGTATCCGAGTCCAGAAAGCACCTGGAACCCGACAGAAAGATTTC 351
29 ..... 29
352 TGGACTCCCGACAGCGGACAGAGAGAGGAGCGGATGAGCGACACACACA 401
30 .....Gly..HisThrG1 33
402 AACACAGAACACACAGCCAGTCCCGAGAGGCCAGTATGAGAGGCCCA 451
33 nThrGlnsnnHisThrAlaSerProArgSerProValMetGlnSerProL 50
452 AAAAGAGAACCCAGCAGCTGAAAGTCGGGATCTACACCTGGGCGACAGAGA 501
50 ysLysLysAsnGlnGlnLeuLysValGlyIleLeuHisLeuGlySerArg 66
502 CAGAGAAGATCAGATACAGTGAATCCAGATCCAGTCCGCGACATGGAAGT 551
67 GlnLysLysIleArgIleGlnLeuArgSerGlnCysAlaThrTrpLysVa 83
552 GATTCGCAAGAGCTGCATCAGTCAAAACACGCGGATTAATCTGGATTGG 601
83 lIleCysLysSerCysIleSerGlnThrProGlyIleAsnLeuAspLeuG 100
602 GTTCGGGGCTCAAGGTGAAGATAATACCTAAAGAGAAACACTGTAAATG 651
100 lysSerGlyValLysValLysIleIleProLysGlnGlnHisCysLysMet 116
652 CCAGAGCAGAGTGAAGAGCAACACCAAGT 681
117 ProGlnAlaGlyGlnGlnGlnProGlnVal 126
```

Sequence	Strd	Orig	ZScore	EScore	Len
SwissProt_39:GCG1_HUMAN	+		138.50	200.41	0.0016
SwissProt_39:SCGA.RABIT	+		114.00	147.61	0.1548
SwissProt_39:SCGB1_HUMAN	+		110.00	157.60	0.2676
SwissProt_39:GGBA.RAT	+		106.00	147.57	0.5743
SwissProt_39:G19P_HUMAN	+		106.00	141.09	0.6161
SwissProt_39:NCR2_MOUSE	+		103.50	124.44	1.111
SwissProt_39:GEBY.RAT	+		103.00	143.35	0.9827
SwissProt_39:CYB_TRYB -			103.00	140.08	1.02
SwissProt_39:DMPI_MOUSE	+		99.50	132.42	1.96
SwissProt_39:NA41_HUMAN	+		99.50	130.94	1.99
SwissProt_39:CA2B_HUMAN	+		98.00	119.78	2.88
SwissProt_39:CYC1EITA	+		97.00	131.51	2.98
SwissProt_39:HNHP_CHICK	+		96.50	133.31	3.18
SwissProt_39:YH35_STERF	+		96.00	130.67	3.55
SwissProt_39:GGE4_HUMAN	+		95.50	133.25	3.51
SwissProt_39:GGE5_HUMAN	+		95.50	139.25	3.51
SwissProt_39:GGE6_HUMAN	+		95.50	139.25	3.51
SwissProt_39:GGE3_HUMAN	+		95.50	139.18	3.51
SwissProt_39:YH16_MOUSE	+		95.00	123.86	4.50
SwissProt_39:YHHA_BORE	+		92.50	126.50	6.59
SwissProt_39:YH63_HSY2H	+		92.50	122.50	6.88
SwissProt_39:GGE2_HUMAN	+		92.00	134.45	7.48
SwissProt_39:YH41_HUMAN	+		92.00	122.32	6.56
SwissProt_39:NA41_CANPA	-		92.00	120.48	7.63
SwissProt_39:SN6C_RAT	-		92.00	116.45	7.97
SwissProt_39:ABL2_HUMAN	+		92.00	114.68	8.12
SwissProt_39:OYX1_HUMAN	+		91.50	124.25	7.95
SwissProt_39:DMPI_RAT	+		89.50	118.71	11.71
SwissProt_39:SN6C_MOUSE	-		89.50	113.22	12.43
SwissProt_39:S24B_ARATH	+		89.00	111.35	13.77
SwissProt_39:GGE7_HUMAN	+		88.50	129.49	12.28
SwissProt_39:GSET_HUMAN	+		88.50	115.90	14.22
SwissProt_39:NCM2_SERCA	+		88.00	122.64	14.35
SwissProt_39:G19P_BOVIN	+		88.00	115.88	15.44
SwissProt_39:CIKE_DROME	+		88.00	109.16	16.61
SwissProt_39:RCO4_HUMAN	+		88.00	108.91	16.65
SwissProt_39:NA41_RAT	-		87.50	114.22	17.06
SwissProt_39:YH1L_EHV	+		87.50	113.36	17.22
SwissProt_39:POLS_RUBRH	-		87.50	109.30	18.00
SwissProt_39:POLS_RUBRH	-		87.50	109.30	18.00
Sequence	Strd	Orig	ZScore	EScore	Len
SwissProt_39:GCG1_HUMAN	+		138.50	200.41	0.0016
SwissProt_39:SCGA.RABIT	+		114.00	147.61	0.1548
SwissProt_39:SCGB1_HUMAN	+		110.00	157.60	0.2676
SwissProt_39:GGBA.RAT	+		106.00	147.57	0.5743
SwissProt_39:G19P_HUMAN	+		106.00	141.09	0.6161
SwissProt_39:NCR2_MOUSE	+		103.50	124.44	1.111
SwissProt_39:GEBY.RAT	+		103.00	143.35	0.9827
SwissProt_39:CYB_TRYB -			103.00	140.08	1.02
SwissProt_39:DMPI_MOUSE	+		99.50	132.42	1.96
SwissProt_39:NA41_HUMAN	+		99.50	130.94	1.99
SwissProt_39:CA2B_HUMAN	+		98.00	119.78	2.88
SwissProt_39:CYC1EITA	+		97.00	131.51	2.98
SwissProt_39:HNHP_CHICK	+		96.50	133.31	3.18
SwissProt_39:YH35_STERF	+		96.00	130.67	3.55
SwissProt_39:GGE4_HUMAN	+		95.50	133.25	3.51
SwissProt_39:GGE5_HUMAN	+		95.50	139.25	3.51
SwissProt_39:GGE6_HUMAN	+		95.50	139.25	3.51
SwissProt_39:GGE3_HUMAN	+		95.50	139.18	3.51
SwissProt_39:YH16_MOUSE	+		95.00	123.86	4.50
SwissProt_39:YHHA_BORE	+		92.50	126.50	6.59
SwissProt_39:YH63_HSY2H	+		92.50	122.50	6.88
SwissProt_39:GGE2_HUMAN	+		92.00	134.45	7.48
SwissProt_39:YH41_HUMAN	+		92.00	122.32	6.56
SwissProt_39:NA41_CANPA	-		92.00	120.48	7.63
SwissProt_39:SN6C_RAT	-		92.00	116.45	7.97
SwissProt_39:ABL2_HUMAN	+		92.00	114.68	8

[illegible]

alignment_block:
US-09-651-563-808 x GRPA_RAT ..

Align seg 1/1 to: GRPA_RAT from: 1 to: 246

```

306 GGGGATCCGAGTCCGAGAACGACCTGGAACCCGAGAAATCTCGA 355
74 G1yserglnlglnglnlglnglnlglnglnlglnglnlglnglnl 90
356 CTTCCGAGAGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 405
90 nglnpProAlaThr.....SerGlys 98
406 CAGAACACAGACGATCCGAGGAGGAGGAGGAGGAGGAGGAGGAG 455
98 erglnlglnglnlglnglnlglnglnlglnglnlglnglnlglng 113
456 GAGAACACGAGCTGAAGTGGGATCTACACCTGGGAGGAGGAGGAG 505
114 ..GluproProAlaThrSerGlySerGlySerGlySerGlySerG 129
506 A.....GAGATCAGATACAGCTGAGATCCAGTGGCGGACA 543
129 uproThrGlnAlaGlnAspGlnGlnProProAlaThrSerGlySer 146
544 TGGAGGTGATCTGCAAGAGCTGATCATGCAACAC.....GGG 584
146 lnglnlglnglnlglnglnlglnglnlglnglnlglnglnlglng 162
585 GATAATCTGATTTGGTTCCGCGCTCAAGAGTGAAGATTAATACCTA 634
163 AspSerAlaGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 179
635 AGGACACACTGTAATGCCAGAGCAGCTGGAAGACCAACCA 678
179 luserProSerSerProGlnAsnSerGlnGlnGlnGlnProGln 193

```

seq_name: SwissProt_39:G19P_HUMAN

seq_documentation_block:

```

ID G19P_HUMAN STANDARD; PRT; 527 AA.
AC P14314;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PROTEIN KINASE C SUBSTRATE, 80 KDA PROTEIN, HEAVY CHAIN (PKCSH)
DE (80K-H PROTEIN).
GN PKCSH OR G19P1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RX MEDLINE=90007553; PubMed=2793184;
RA Sakai K., Masamichi H., Minoshima S., Kudoh J., Fukuyama R.,
RA Shimizu N.,
RT "Isolation of cDNAs encoding a substrate for protein kinase C;
RT nucleotide sequence and chromosomal mapping of the gene for a human
RT 80K protein."
RL Genomics 5:309-315(1989).
RL [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97196775; PubMed=9043864;
RA Opoit R.A., Terwindt G.M., Vergouwe M.N., van Eljk R.,
RA Mohrenweiser H., Iltt M., Hofker M.H., Haan J., Ferrari M.D.,
RA Frants R.R.;
RT "A 3-Mb region for the familial hemiplegic migraine locus on 19p13.1-
RT p13.2: exclusion of PKCSH as a candidate gene."
RL Eur. J. Hum. Genet. 4:321-328(1996).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

```

CC EMBL: J03075; AAA52493.1; -.
DR EMBL: U50326; AAA98668.1; JOINED.
DR EMBL: U50317; AAA98668.1; JOINED.
DR EMBL: U50318; AAA98668.1; JOINED.
DR EMBL: U50319; AAA98668.1; JOINED.
DR EMBL: U50320; AAA98668.1; JOINED.
DR EMBL: U50321; AAA98668.1; JOINED.
DR EMBL: U50322; AAA98668.1; JOINED.
DR EMBL: U50323; AAA98668.1; JOINED.
DR EMBL: U50324; AAA98668.1; JOINED.
DR EMBL: U50325; AAA98668.1; JOINED.
DR PIR: A32469; A32469.
DR HSSP: P06705; 1AUI.
DR MIM: 177060; -.
DR InterPro: IPR000886; -.
DR InterPro: PS00014; ER_TARGET; 1.
DR PROSITE: PS00018; EF_HAND; 1.
KW Phosphorylation.
FT DOMAIN 313 336 GLU-RICH (ACIDIC).
FT MOD_RES 89 89 PHOSPHORYLATION (BY PKC) (POTENTIAL).
FT MOD_RES 382 382 PHOSPHORYLATION (BY PKC) (POTENTIAL).
FT MOD_RES 389 389 PHOSPHORYLATION (BY PKC) (POTENTIAL).
FT MOD_RES 433 433 PHOSPHORYLATION (BY PKC) (POTENTIAL).
SQ SEQUENCE 527 AA; 59296 MW; 2BB2F9AC14E3BC94 CRC64;

```

alignment_scores:
Quality: 106.00 Length: 212
Ratio: 1.000 Gaps: 10
Percent Similarity: 50.000 Percent Identity: 24.528

alignment_block:

US-09-651-563-808 x G19P_HUMAN ..

Align seg 1/1 to: G19P_HUMAN from: 1 to: 527

```

52 TCCGCTACTGACAGACCGCGGTAAGTCCACAGGACGATCCAACTGGAG 101
230 SerValThrGlnLeu...GlnThrHisProGlnLeuAspThrAspGlyAs 245
102 TTGAAGTGTGAGTACAGTGAA..... 123
245 GAGAACACGAGCGCTCCGAGGAGGAGGAGGAGGAGGAGGAGGAG 262
262 lnrThrAspAlaThrSerPheThrAspArgValThrPheAlaAla.IleArg 278
172 GAGAGAGCCCTGCAAGTCGTCGTCCTGATCGGAGTGGCAGCCCATGG 221
278 pLysTyrArgSerGln.....AlaLeuProThrA 288
222 ACCCTCTTGTCTGTCACGCGCATTAAGTAGGAGG...AAGAGGGCCGA 268
288 spleuProAlaProSerAlaProAspLeuThrGlnProLysGlnGln 304
269 GAGGTGAGGGGCTCAGGCGAAGCTGGGGTCTGTTGGGGTATCCGAGT 318
304 ..... 304
319 CCAGAGAACACCTGGAACCCGAGACAGAGATTCGACTCCACAGACGG 368
305 PropProValProSerSerProThrGlnGlnGln.....GlnGlnG 318
369 ACCAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 418

```


alignment_scores:

Quality: 103.00 Length: 139
 Ratio: 1.304 Gaps: 7
 Percent Similarity: 56.835 Percent Identity: 28.058

alignment_block:

US-09-651-808 x GRPB-RAT ..

Align seg 1/1 to: GRPB-RAT from: 1 to: 247

```

306 GGGGTATCCAGTCCAGAGACACCTGGAAACCCGACAGAGAAATTCTGCA 355
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
74 GlySerGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 90
356 CTCCCGGAGGGGACGAGAGAGAGGAGGATGATGACACACACAAACA 405
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
90 nGluProAlaThr.....SerGlyS 98
406 CAGAACACACAGACGACCTCCAGAGCCAGT.....AATGAGAG 446
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
98 erGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 114
447 CCC.....CAAAAGAGAGACGAC 466
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
115 ProProAlaThrSerGlySerGlySerGluGluGluGluGluGluGlu 131
467 A...GCTGAAGTGGGATCCTACACCTGGGACAGACAGAGAGAAATC 513
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
131 rGlnAlaGluGluGluGluGluGluGluGluGluGluGluGluGlu 148
514 AGGATACAGCTGAGATCCAGTGGCGACATGGAAGTGATCTGCAAGAG 563
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
148 lngGluGluGlu.....SerThrGlnAlaGluGluGluGluGlu 160
564 CTGCATCAGTCAACACCGGGGATTAATCTGGATTGGTCCGGCTCA 613
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
161 .....ProSerAspSerAlaGlyGluGluGluGluGluGluGlu 172
614 AGTGAGATATATACCTTAAGAGACACATGTAATAATCCAGAGACAGT 663
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
172 nProGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 188
664 GAGAGCAACCAACAA 678
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
189 GlnGluGluGluGlu 193

seq_name: SwissProt_39:CYB_TRYB
seq_documentation_block:
ID CYB_TRYB STANDARD; PRT; 363 AA.
AC P00164;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CYTOCHROME B.
OS COB OR CYTB.
OS Trypanosoma brucei brucei.
OC Mitochondrion.
OC Eukaryota; Euzoenzoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5702;
RN [1]
RP SEQUENCE FROM N.A. (KINETOPLAST).
RX MEDLINE=84041494; Pubmed=6314266;
RA Beane R., de Vries B.F., van den Burg J., Klaver B.:
RT "The nucleotide sequence of a segment of Trypanosoma brucei
RT mitochondrial maxi-circle DNA that contains the gene for
RT apocytochrome b and some unusual unassigned reading frames.";
RL Nucleic Acids Res. 11:6925-6941(1983).
CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS.

```

```

CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN.
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RISKE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X00017; CAA24915.1; -.
DR PIR: A00160; CBUTB.
DR InterPro: IPR000179; -.
DR Pfam: PF00032; cytochrome_b_N; 1.
DR Pfam: PF00033; cytochrome_b_N; 1.
DR PROSITE: PS00192; CYTOCHROME_B_HEME; 1.
DR PROSITE: PS00193; CYTOCHROME_B_OO; 1.
KW Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
KW Heme; Kinetoplast.
FT METAL 74 74 IRON 1 (HEME B562 AXIAL LIGAND).
FT METAL 88 88 IRON 2 (HEME B566 AXIAL LIGAND).
FT METAL 175 175 IRON 2 (HEME B562 AXIAL LIGAND).
FT METAL 189 189 IRON 1 (HEME B566 AXIAL LIGAND).
SQ SEQUENCE 363 AA; 43455 MW; 2A37CD827E55923B CRC64;

```

alignment_scores:

Quality: 103.00 Length: 191
 Ratio: 1.288 Gaps: 12
 Percent Similarity: 41.885 Percent Identity: 23.560

alignment_block:

US-09-651-563-808/rev x CYB_TRYB ..

Align seg 1/1 to: CYB_TRYB from: 1 to: 363

```

705 GTTGTTTACCTGCTCTCATTTAACTTGTTGTTGCTCTCACCTGCT 656
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
37 IleCysGlyValCysLeu.....AlaTrpLeuPhePheSerCysSh 50
655 C.....TGGCAATTTACAGTGTCTCTTAACTGATATATCTTC 618
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
50 eileCysSerAsnTrpTrpPheValLeuPheLeuTrpAspPhe..... 64
617 ACCTTGACCGCGGACCAATCCAGATTATCCCGGCTTTGACTGAT 568
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
65 .....AspLeuGlyPheValIleArgSerValHisIleCysPheThrSer 79
567 GCAGCTCTTCAGATCACCTTC.....CATGCGCGCAGTGGATCTCAGCTATCTG 513
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
80 Leu.LeuTyrLeuLeuLeuTyrIleHisIlePheLysSerIleThrLeu 96
545 .....CATGCGCGCAGTGGATCTCAGCTATCTG 513
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
96 lelleLeuPheAspTrpHisIleLeuValTrpPheIleGlyPheIleLeu 112
512 ...ATCTTCTTCTGTCTGCTGCGCCAGCTAGATCCGACTTTCAGCTG 466
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
113 PheValPheIleIleIleIleAlaPheIleGlyTyrValLeuProCysT 129
465 CTGGTCTCTCTTTTGGGCG.....CATGCGCGCAGTGGATCTCAGCTATCTG 446
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
129 hrnelelelelelelelelelelelelelelelelelelelelele 145
445 TCTCCATTACTGGGCTCTG.....GACTGCGCTGTGGTCTGTGTT 402
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
146 ValProIleleuGlyIleTrpLeuGlyCysTyrTrpIleTrpGlySerLuh 162
401 TGTG.....TGTGCGCTATGCGCGGTCCTCTCTGCTGCTC 367

```

```

||||:||||:||||:||||:
162 eileasnasphephtleuleulysleuhsivalleuhsvalleuleup 179
366 CG..... 365
179 ropheleuleullelleuleulleuhsleuphecysleuhsityr 195
364 rctgggagctccagaaatctctgt.....CG 339
196 phemeiserersaplaaphecysasparphealaphetyrcysgluar 212
338 gggtccagagctctctg 320
212 gleuserphecysmetirp 218

seq_name: SwissProt_39:DMP1_MOUSE

seq_documentation_block:
ID DMP1_MOUSE STANDARD; PRT; 503 AA.
AC 05188;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DENTIN MATRIX ACIDIC PHOSPHOPROTEIN 1 PRECURSOR (DENTIN MATRIX
DE PROTEIN-1) (DMP-1) (AG1).
GN DMP1 OR DMP
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SWISS WEBSTER; TISSUE=Molar;
RX MEDLINE=9818425; PubMed=9525343;
RA McDougall M., Gu I.T., Luan X., Simmons D., Chen J.;
RT "Identification of a novel isoform of mouse dentin matrix protein 1:
RT spatial expression in mineralized tissues.";
RL J. Bone Miner. Res. 13:422-431(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SVJ; TISSUE=Spleen;
RA Feng J.Q., Trianedes K., Luan X., McDougall M.;
RT "Study of murine Dmp-1 gene function and regulation.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COULD BE INVOLVED IN THE MINERALIZATION OF EXTRACELLULAR
CC MATRIX.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN TOOTH PARTICULARLY IN
CC ODONTOBLAST, AMELOBLAST AND CEMENTOBLAST. ALSO EXPRESSED IN BONE
CC PARTICULARLY IN OSTEOBLAST.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U65020; AAB93764.1; -.
DR EMBL: AJ242625; CAB59629.1; -.
DR MGD: MGI:94910; Dmp1.
KW Extracellular matrix; Signal.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 503 DENTIN MATRIX ACIDIC PHOSPHOPROTEIN 1.
FT DOMAIN 41 44 POLY-PRO.
FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 457 457 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SITE 350 352 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 350 352 D -> G (IN REF. 2).
FT CONFLICT 35 35 H -> D (IN REF. 2).
FT CONFLICT 67 67 D -> E (IN REF. 2).
FT CONFLICT 99 99 G -> Y (IN REF. 2).
FT CONFLICT 116 116

```

```

FT CONFLICT 137 137 A -> T (IN REF. 2).
SQ SEQUENCE 503 AA; 54000 MW; 9E8AF9F2729F113A CRC64;

alignment_scores:
  Quality: 99.50      Length: 281
  Ratio: 0.809      Gaps: 12
  Percent Similarity: 43.772      Percent Identity: 23.132

alignment_block:
US-09-651-563-808 x DMP1_MOUSE ..

Align seq 1/1 to: DMP1_MOUSE from: 1 to: 503

22 GACTGGGCTCCGAGGCTGATTCCTCCGCTACTGACACGCGG 71
|||||
104 Aspergly.....AspaPTThrPheGlyAspGluAspAsnG 116
72 GGTAGGTCACAGGACGATCCAACTGGAGTGAAGTGTGAGAGAGT 121
|||||
116 yLeuGlyProGluGluGlyGlnTrpGlyGlyProSerTysLeuAspSera 133
122 AAGAGAACGACGAGGCTCCGAGGAGTGTGTGTCAGTACGACGAGT 171
133 spGluAspSera.....AspThrThrGlnSer 142
172 GAGAGGCGCTCGAAGTCGTGCTCCTCATGCGGTGCCAGCCCATG 221
|||||
143 SerGluAspSerThrSerGlnGluAsnSeraLagAspThrProSera 159
222 ACCTCTT.....GTCTGTCAGCGCCCATACTA 250
159 spSerTysAspGluAspSerGluAspAspAlaHisSerThrProAspAla 175
251 GGG.....AGGAGAGCGCCGACG 270
176 GLyAspSeraLagInHisSerGluSerGluGluArgValGlyGly 192
271 AGTGAAGGCGCTCGAAGCTGGGT..... 298
192 ySerGlyGly...GlnSerSerHisGlyAspGlySerGluPheAspSg 208
298 ..... 298
208 LuGlyMetGlnSerAspAspProGluSerThrArgSerAspArgGlyHis 224
299 .....GCTGTGGGGTATCCAGTCCAGAACCTGGA..... 334
225 AlaArgMetSerSerLagGlyLeaArgSerGluGluSerTysGlyAspAr 241
335 .....ACCCGACAGAGATTCGTGACGCCGACG..... 364
241 gGluProThrSerThrGlnAspSerAspAspSerGlnSerValGluPheS 258
364 ..... 364
258 erSerArgTysSerPheArgSerArgValSerGluGluAspTyrArg 274
365 .....CGGACACGAGAGAGGAGGAGCATGAG 390
275 GlyGluLeuThrAspSerAsnSerArgGluTrnGlnSerAspSerThrG 291
391 CGACACA...CACAAACACAGAACACACAGACGATCCGAGAGCCAGT 437
291 uAspThrAlaSerTysGluGluSerArgSerGluSerGlnGlu.....Asp 306
438 AATGAGAGCCCAAAAGAGAACACAGCAGCTGAAGTGGAGCTTAC 487
307 ThrAlaGluSerGlnSerGlnGluAspSerProGluGlyGluAspProse 323
488 ACCTGGGACGACGACGAGAGAGATGATACAGTGTGATCCAGTGC 537
323 rSerGluSerSerGluGluAlaGlyGluProSerGlnGluSerSerSerg 340

```



```

ID seq documentation_block: PRT: 1736 AA.
1D CA2B_HUMAN STANDARD
AC P13942: Q13273: Q13271: Q13272: Q0751,
AD 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE COLLAGEN ALPHA 2(XI) CHAIN PRECURSOR.
GN COL1A2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN NM_001111
RP SEQUENCE FROM N.A.
RX MEDLINE=96032717; Pubmed=7559422;
RA Vioristo M.M., Fihlaajmaa T., Vandenberg P., Prockop D.J.,
RA Ala-Kokko L.;
RT "The human COL1A2 gene structure indicates that the gene has not
RT evolved with the genes for the major fibrillar collagens."
RL J. Biol. Chem. 270:22873-22881(1995).
RN [2]
RP SEQUENCE OF 59-807 FROM N.A.
RC TISSUE=Cartilage;
RX MEDLINE=93314796; Pubmed=8325374;
RA Zhidkova N.I., Brewton R.G., Mayne R.;
RT "Molecular cloning of PARP (proline/arginine-rich protein) from human
RT cartilage and subsequent demonstration that PARP is a fragment of the
RT NH2-terminal domain of the collagen alpha 2(XI) chain."
RL FEBS Lett. 326:25-28(1993).
RN [3]
RP SEQUENCE OF 730-1690 FROM N.A.
RX MEDLINE=89340485; Pubmed=2760050;
RA Kimura T., Cheah K.S.E., Chan S.D.H., Lui V.C.H., Mettel M.-G.,
RA van der Rest M., Oio K., Solomon E., Nishimura Y., Olsen B.R.;
RT "The human alpha 2(XI) collagen (COL1A2) chain. Molecular cloning of
RT cDNA and genomic DNA reveals characteristics of a fibrillar collagen
RT with differences in genomic organization."
RL J. Biol. Chem. 264:13910-13916(1989).
RN [4]
RP ALTERNATIVE SPLICING.
RX MEDLINE=95238468; Pubmed=7721876;
RA Zhidkova N.I., Justice S.K., Mayne R.;
RT "Alternative mRNA processing occurs in the variable region of the
RT pro-alpha 1(XI) and pro-alpha 2(XI) collagen chains."
RL J. Biol. Chem. 270:9486-9493(1995).
RN [5]
RP REVIEW ON VARIANTS.
RX MEDLINE=97255959; Pubmed=9101290;
RA Kuivaniemi H., Tromp G., Prockop D.J.;
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT associated collagen (type IX), and network-forming collagen (type X)
RT cause a spectrum of diseases of bone, cartilage, and blood vessels."
RL Hum. Mutat. 9:300-315(1997).
RN [6]
RP VARIANT ARG-661.
RX MEDLINE=95163096; Pubmed=7859284;
RA Viikula M., Mairman E.C.M., Lui V.C.H., Zhidkova N.I., Tiller G.E.,
RA Goldring M.B., Van Beersum S.E.C., de Waal Malefijt M.C.,
RA van den Hoogen F.H.J., Rogers H.-H., Mayne R., Cheah K.S.E.,
RA Olsen B.R., Warman M.L., Brunner H.G.;
RT "Autosomal dominant and recessive osteochondrodysplasias associated
RT with the COL1A2 locus."
RL Cell 80:431-437(1995).
RN [7]
RP -I- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN FIBRILLOGENESIS BY
CC CONTROLLING LATERAL GROWTH OF COLLAGEN II FIBRILS.
CC -I- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(XI),
CC ALPHA 2(XI), AND ALPHA 3(XI). ALPHA 3(XI) IS A POST-TRANSLATIONAL
CC MODIFICATION OF ALPHA 1(II). ALPHA 1(V) CAN ALSO BE FOUND INSTEAD
CC OF ALPHA 3(XI)-1(II).
CC -I- ALTERNATIVE PRODUCTS: ALTERNATIVE FORMS OF THIS COLLAGEN ARE
CC PRODUCED BY ALTERNATIVE SPLICING. THESE FORMS DIFFER IN THE
CC PRESENCE OR ABSENCE OF THREE SEPARATE REGIONS IN THE N-TERMINAL

```

alignment_scores: Quality: 98.00 Length: 183
 Ratio: 1.167 Gaps: 10
 Percent Similarity: 45.902 Percent Identity: 26.230

alignment_block:

US-09-651-563-808 x CA2B_HUMAN ..

Align seg 1/1 to: CA2B_HUMAN from: 1 to: 1736

```

180 CCTGCAAGTCGTGCTCCCTCATGCGGTGCGACGCCACGACCTTCTT 229
176 ProArgSerAlaArgPro.....ValLe 183
230 GTCCTGTCAGCGGCCATACAGGAGGAGGAGG.....GCCGAGG 270
183 uasPTThHisGlyValIleIlePheGlyAlaArgIleLeuAspLcyluv 200
271 AGTGGAGGGGCTCA...GGCGAGCTGGGCTGTGGGGGTATCCGAG 317
200 alPheGluGlyAspValGlnGlnLeuAlaIleValProGlyValGlnAla 216
318 TCCGAGAGGACCTGGAACCCCGACACAGATCTGTGACTCCCGACGCG 367
217 AlAtyGlnSerGly.....GlnLysGlnLysGlnLysGlnLysGln 231
368 GACGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 416
231 yGlnArgGlu.....ArgProGlnAsnGlnGlnProHisA 243
417 .....AGCCAGTCCCGAGGAGGCCAGTAAT...GGAGAGCCCAA 452
243 rgAlaGlnArgSerProGlnGlnGlnProSerArgLysHisArgProGln 259
453 AAAGAGAGACGACGACGAGTAAT..... 476
260 AsnGlnGlnProGlnSerGlnProThGlnSerLeuTyrTyrArgI 276
477 .....CGCG 480
276 uProProTyrTyrAspValMetThrThrGlyThrThrProAspTyrGlnA 293
481 ATCTACACCTGGGCGACGACGACGACGACGACGACGACGACGACGAC 530
293 sPrProThrProGlyGlnGlnGlnLysLeuGlnSerLeuLeuPro 309
531 CCAGGCGGCGACATGAGGAGTGCATGAGGAGTGCATGAGTGCATGAG 580
310 ProLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 323
581 CGGGGATTAATCTGATTGTTGGTCCGCGCTCAAGTGAAGTAATACG 629
323 rAlaAspArg.....PheGlnAlaGlnGlnGlnGlnGlnGlnGlnGln 337

```

seq_name: SwissProt_39:CYB_LEITA

seq_documentation_block:

ID CYB_LEITA STANDARD: PRT: 371 AA.
 AC P14548:
 DT 01-JUN-1990 (Rel. 13, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE CYTOCHROME B.
 GN COB OR CYTB OR CYB.
 OS Leishmania tarentolae (Sauroleishmania tarentolae).
 OC Mitochondrion.
 NC Eukaryota; Euzoenzoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 NCBI_TaxId=5689;
 RN [1]
 RP SEQUENCE FROM N.A., AND RNA EDITING.
 RX MEDLINE=88210465; PubMed=2452696;
 RA Shaw J.M., Feagin J.E., Stuart K.D., Simpson L.;
 RT "Editing of kinetoplastid mitochondrial mRNAs by uridine addition and

RT deletion generates conserved amino acid sequences and AUG initiation
 RT codons.";
 RL Cell 53:401-411(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85297791; PubMed=2994021;
 RA Simpson A.G., Necklemann N., la Cruz V.F., Mulich M.L., Simpson L.;
 RT "Mapping and 5' end determination of kinetoplast maxicircle gene
 RT transcripts from Leishmania tarentolae.";
 RL Nucleic Acids Res. 13:5977-5993(1985).
 RN [3]
 RP SEQUENCE OF 1-48 FROM N.A.
 RX MEDLINE=88124876; PubMed=2448777;
 RA Feagin J.E., Shaw J.M., Simpson L., Stuart K.;
 RT "Creation of AUG initiation codons by addition of uridines within
 RT cytochrome b transcripts of kinetoplastids.";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:539-543(1988).
 RN [4]
 RP SEQUENCE OF 21-371 FROM N.A.
 RX MEDLINE=85079995; PubMed=6096360;
 RA de la Cruz V.F., Necklemann N., Simpson L.;
 RT "Sequences of six genes and several open reading frames in the
 RT kinetoplast maxicircle DNA of Leishmania tarentolae.";
 RL J. Biol. Chem. 259:15136-15147(1984).
 CC -i- FUNCTION: COMPONENT OF THE URIDINOL-CYTOCHROME C REDUCTASE
 CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
 CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
 CC COUPLED TO ATP SYNTHESIS.
 CC -i- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
 CC BOUND TO THE PROTEIN.
 CC -i- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
 CC CYTOCHROME C1 AND THE RIESKE PROTEIN.
 CC -i- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
 CC -i- CAUTION: THE GENOMIC DERIVED SEQUENCE DIFFERS FROM THAT SHOWN AS
 CC IT IS MODIFIED BY EXTENSIVE RNA EDITING.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: L07542; AAA31879.1; ALT_SEQ.
 DR EMBL: M10126; ?; NOT_ANNOTATED_CDS.
 DR EMBL: M19065; AAA31878.1; ?.
 DR PIR: H22848; H22848.
 DR PIR: A28118; A28118.
 DR InterPro: IPR000179; ?.
 DR Pfam: PF00032; cytochrome_b_C1; 1.
 DR Pfam: PF00033; cytochrome_b_N; 1.
 DR PROSITE: PS00192; CYTOCHROME_B_HEME; 1.
 DR PROSITE: PS00193; CYTOCHROME_B_OO; 1.
 KW Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
 KW Heme.
 FT METAL 82 82 IRON 1 (HEME B562 AXIAL LIGAND).
 FT METAL 96 96 IRON 2 (HEME B566 AXIAL LIGAND).
 FT METAL 183 183 IRON 1 (HEME B562 AXIAL LIGAND).
 FT METAL 197 197 IRON 2 (HEME B566 AXIAL LIGAND).
 SQ SEQUENCE 371 AA; 44555 MW; 0003BD11538EA75B CRC64;

alignment_scores:

Quality: 97.00 Length: 152
 Ratio: 1.347 Gaps: 10
 Percent Similarity: 47.368 Percent Identity: 23.684

alignment_block:

US-09-651-563-808/rev x CYB_LEITA ..

Align seg 1/1 to: CYB_LEITA from: 1 to: 371

Fri Jun 29 10:17:49 2001

us-09-651-563-808.rsp

Page 14

seq_name: sp.bacteria:Q9L8W2

seq_documentation_block:

ID Q9L8W2 PRELIMINARY; PRT; 355 AA.

AC Q9L8W2;

DT 01-OCT-2000 (TReMBLe1.15, Created)

DT 01-OCT-2000 (TReMBLe1.15, Last sequence update)

DT 01-OCT-2000 (TReMBLe1.15, Last annotation update)

DE 87-KDA SURFACE LIPOPROTEIN PRECURSOR (FRAGMENT).

GN VLPAB7.

OS Mycoplasma hyorhinis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;

OX NCBI_TaxID=2100;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-SK76;

RX MEDLINE=20138163; PubMed=10671459;

RA Citti C., Watson-McKown R., Drosse M., Wise K.S.;

RT "Gene Families Encoding Phase- and Size-Variable Surface Lipoproteins

RT of Mycoplasma hyorhinis."

RL J. Bacteriol. 182:1356-1363(2000).

DR EMBL; AF193878; AAF36546.1; -

KM Lipoprotein.

FT LIPOTERIN. 1 1

SEQUENCE 355 AA; 35197 MW; D3D638067B2E117F CRC64;

```
alignment_scores:
  quality: 104.50
  length: 241
```

```

Ratio: 0.760           Gaps: 11
Percent Similarity: 55.602      Percent Identity: 26.141

alignment_block:
US-09-651-563-808 x Q9J8W2 ..

Align seg 1/1 to: Q9J8W2 from: 1 to: 355

```

```
alignment_block:
```

US-09-651-563-808 x Q9L8W2

Align seg 1/1 to: Q9L8W2 from: 1 to: 355

[illegible]

seq_name: sp_bacteria:Q9L8V9

```
seq_documentation_block:
ID      Q9L8V9      PRELIMINARY;      PRT;      384 AA
AC      Q9L8V9;
```

DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
 DE 87-KDA SURFACE LIPOPROTEIN PRECURSOR.
 GN VLPAB7.
 OS Mycoplasma hyorhinis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 OC Mycoplasmataceae; Mycoplasma.
 OC NCBI_TaxID=2100;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SK76;
 RA MEDLINE=20138163; PubMed=10671459;
 RA Citi C., Watson-McKown R., Drosse M., Wise K.S.;
 RT "Gene Families Encoding Phase- and Size-Variable Surface Lipoproteins
 of Mycoplasma hyorhinis";
 RL J. Bacteriol. 182:1356-1363(2000).
 DR EMBL: AF193880; AAF36549.1; -.
 KW Signal; Lipoprotein.
 FT SIGNAL 1 29 POTENTIAL.
 FT CHAIN 30 384 87-KDA SURFACE LIPOPROTEIN.
 SQ SEQUENCE 384 AA; 38226 MW; 04A1F0A98E37B79C CRC64;

alignment_scores:
 Quality: 104.50 Length: 241
 Ratio: 0.780 Gaps: 11
 Percent Similarity: 55.602 Percent Identity: 26.141

alignment_block:
 US-09-651-563-808 x Q9L8V9 ..

Align seg 1/1 to: Q9L8V9 from: 1 to: 384

```

16 GCCGGCGACTCGGGTCCCGAGCTCGAFTCTTCTCCGCTACTGAGAC 65
   ::::::::::::::::::::
111 SerGlySerAsnSerGlySerGlyMetAsnSerGlyThrGluAsn 127
66 ACGGGGGAGTGTCCACGAGCATCTGGAGTGAAGTGAAGTGAAGG 115
   ::::::::::::::::::::
127 rGlnGlnSerGluAlaProGlyThrAsnThrGlyAsnThrThrSerg 144
   ::::::::::::::::::::
116 AGAGTAGAGGAA.....CCAGCAGCTTCCGAGGAGGT 150
   ::::::::::::::::::::
144 IuserAsnSerGlySerThrGlySerGluAlaGlyThrThrAsn 160
   ::::::::::::::::::::
151 GRTGTGTGAGTCTAGAGTGAAGGCCCTC.....GAAGTGTGCT 194
   ::::::::::::::::::::
161 ThrGlySerGlySerAsnSerGlySerGlyMetAsnSerGlyThr 177
   ::::::::::::::::::::
195 CCCTTCATCGCGTGGCCACCGCATGCTTGTCTGTCGTCACGAGCA 244
   ::::::::::::::::::::
177 uAsnThrGlnSerGluAlaProGlyThrLys.....ThrGlu 191
   ::::::::::::::::::::
245 TAAGTAGAGGAGGAGGAGGCCGAGA.....GTGAGAGGCTCAGGC 288
   ::::::::::::::::::::
191 snThrGlnSerGluAlaPro. GlyThrLysThrGluAsnThrGln 207
289 AAGCTGGGGTCTGTGGGGTATCCGAGTCCAGAGCAGCTGAGACC 338
   ::::::::::::::::::::
207 nser.....GluAlaProGlyThr 214
   ::::::::::::::::::::
339 CCACAGAGAT.....TCTGACTCCCGAGCGGAGCAGGAGG 379
   ::::::::::::::::::::
214 snThrGlyAsnLysThrThrSer. GluSerAsnSerGlySerThr 230
380 GACGGCATAGGAGACACACAAACACAGACACAGCCAGTCCAGG 429
   ::::::::::::::::::::
230 ySerGluAlaGlyThrThrAsnThrGlySerGlySerAsnSerGlu 247
430 ACCCCGATATGAGAGCCCAAAAAGAAAGAACACGACGTAAGTGG 479
   ::::::::::::::::::::
247 ergLysMetAsnSerGlyThrGluAsnThrGlnSerGluAla... 262

```

480 GATCTACACTGGGACAGACAGAAAGATCAGATACAGTGAAT 529
 ||||| :::::::::::::: ::::: :::::
 263ProGlyThrLysThrGluAsnThrGln...SerGlu 275
 530 CCCAGTGGCGACATGAGGTATCTGCA...GACTGCATCAGCAA 576
 ||| ::::: ||| ||| :::::
 275 aProGlyThrLysThrGluAsnThrGlnSerGluAlaLeuGlyThr 292
 577 ACACGGGAGATTAATCTGATTTGGTTCGCGCTCAAGGTGAATAT 626
 ||||| :::::::::::::: ::::: :::::
 292 snThrGlyAsnLysThrThrSerLysSerAsnSerGlySerThr 308
 627 ACCTAAGAGGACATGTAATATCCAGAGCAGTGAAGCAACAC 676
 ::| ::::: ||| ::::: :::::
 309 SerGluAlaGlyThrThrAsnThrGlySerGlySerAsnSerGlu 325
 677 AAGTTAATGAGACAGCAAG 695
 ::::::::::::::::::::
 325 rGlyMetAsnSerGluLys 331

seq_name: sp.rodent:Q9ET15

seq_documentation_block:
 ID Q9ET15; PRELIMINARY; PRT; 538 AA.

AC Q9ET15;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DE 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE RETINITIS PIGMENTOSA GTPASE REGULATOR (FRAGMENT).
 GN RPR.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Verwoort R., Lennon A., Bird A.C., Tullioch B., Axion R., Miano M.G.,
 RA Weidol A., Mellinger T., Ciccodicola A., Wright A.F.,
 RT "Mutational hot spot within a new RPR exon in X-linked retinitis
 pigmentosa";
 RT Nat. Genet. 0:0-0(2000).
 RL EMBL: AF286473; AAG00552.1; -.
 DR EMBL: AF286473; AAG00552.1; -.
 FT NON_TER 1 1
 FT SEQUENCE 538 AA; 60333 MW; 549F3E09243FDE0 CRC64;

alignment_scores:
 Quality: 104.50 Length: 228
 Ratio: 0.893 Gaps: 11
 Percent Similarity: 51.316 Percent Identity: 25.877

alignment_block:
 US-09-651-563-808 x Q9ET15 ..

Align seg 1/1 to: Q9ET15 from: 1 to: 538

```

109 GTGAGTAGAGTGAAGAGCAACAGAGCTTCCGAGG.....GTGT 152
   ::::::::::::::::::::
   1 IleProGlnGlnGlnGlnGlyProGluAspSerGluLysAsnValVal 17
153 GTGTGAGTGAAGTGAAGAGGCC...CTGCAAGTGTGCTGCC 199
   ::::::::::::::::::::
   17 IglGlnValValGlnAlaGlnLysGluAsnLeuGlnPheGlnGlyAsp 34
200 TCATGCGGTGCCAGCCCATGACCTTCTGCTGCTCAGC..... 240
   ::::::::::::::::::::
34 rGlyGlnAlaLysAlaGlnAlaProSerAspValIleThrGluLysGlu 50
241 .....GCCATACTAGAGGAGGAGGCCCGAGAGAGGAGGAGG 283
   ::::::::::::::::::::
51 ValSerGluSerGluArgGluSerGlyGlyGluArgGluAspArgSer 67

```

```

284 AGCGAAGCTGGGGTCTGTG..... 305
    |||:||||:
67 uGlyAspGlyAspGlnIleCysGluLysValSerLeuGluThrGluHisL 84
306 ..... GGGGTATCGAGTCCGAGAACGACCTGGAAACCCCGA 341
    |||:||||:
84 euGlnArgAlaGlnGlyLysGlnIleArgLysLysGlyLysAspLysArg 100
342 CAGAG...ATTCTGCATCCCGAC.....GGGACGAG 373
    |||:||||:
101 AlArgGlySLeuAspMetLysGluArgGluGluAspLysGlyTrpGlu 117
374 GAGAGGAGCGCATGAGCGACACAAACAGAACACACAGCCAGT 423
    |||:||||:
117 uLysGlySerGluGlyLysAspLysMetLysArgGluGlyAsnGlnG 134
424 CC.....CAGAGCCCGATGAGAGCCCGCAAAA 455
    |||:||||:
134 LuLysArgLysLysGluMetGluIleArgAspAlaGlyAspGluArgSer 150
456 GAAGAACACAGCTGAAAGTCCGATCTACACCTGGCGACAGACA... 503
    |||:||||:
151 GluGluGluLysGluGluGluGluGluGluGluGluGluGluGluGlu 167
504 .....GAAGAAGTACAGATACAGTCCAGTCCGCGGACATGG 546
    |||:||||:
167 uGlyGlyGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 184
547 AAGGTATCTGCAAGCTGATCATGCAACACCGGGATTAATCTGGA 596
    |||:||||:
184 LuGlyAspArgGlnGlu.....LysGluGlyArgArgGluGly 196
597 TTGGGTCTCCGGCTCAAGTGAAGATATACCTAAAGAGCAACTGTGA 646
    |||:||||:
197 LysGlyArgGlnGluAspGlyArgGluGlyTrpLysGlu..... 209
647 AATGCCAGAACGATGAAGACCAACCA 678
210 .....GlyGluGluGluGluGlu 215
seq_name: sp.bacteria:Q9RV01

seq_documentation_block:
ID Q9RV01 PRELIMINARY; PRT; 319 AA.
AC Q9RV01;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 33.1 KDA PROTEIN.
GN DRI229.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RI;
RA MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.D., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans RI.";
RL Science 286:1571-1577(1999).
DR EMBL; AF001971; AAF10810.1; -.
KW Hypothetical protein.
SO SEQUENCE 319 AA; 33082 MW; 7F0229C5DAAA2B61 CRC64;

```

```

alignment_scores:
    Quality: 104.00      Length: 159
    Ratio: 1.182        Gaps: 12
    Percent Similarity: 55.346    Percent Identity: 30.818

```

alignment_block:

US-09-651-563-808 x Q9RV01 ..

Align seg 1/1 to: Q9RV01 from: 1 to: 319

```

19 GCGAGCTCGGTCCTGAGAGTGTGATCTTCTCCGCTACGACAGC 68
    |||:|||||
159 GlySerSerGlyProSerSerGlyCysSerAlaAsnThrThrArgAla 175
69 GCGGAGTAGTCCACAGCGCATGATCCAACTGGAGTTGAAGTGTGAGTGA 118
    |||:|||||
175 gHisAlaProAlaArgSerSerProLeuAlaSer.....T 187
119 GTGAGAGAACCCAGCAGCTTCGGA.....GG 147
    |||:|||||
187 hGlyHisSerProAlaAlaSerGlyTrpGlyArgSerProSerArgAs 203
148 GTTGTGTGTGCTAGTACTGAGTGAAGAGCCCTGGAAGTGTGCTGCC 197
203 nTrpCysGly.....AlaAlaArgProSerValSerSerGlyA 216
198 TCTCATGCTGTCGCCAGCCATGACCTTGTCTGCTCAGCGCCATTA 247
    |||:|||||
216 rGlyAlaThrSerThrArgCysTrp.ProAlaAlaSerAlaAsnSerVal 232
248 CTAGGAGGAGAGAGGCGCCAGAGTGAAGGCTCAGGCGCAAGCTGGGG 297
    |||:|||||
232 rOlySerProSerSerProArgSerValArgAlaGlyThrArgTrpGly 248
298 TGCTGT...TGG.....GGGTATCGG...AGTCCAGAGAGCACC..... 330
249 CysCysThrThrThrProGlyAlaProAlaAlaProAlaGluThrThr 265
331 .....TGAGCCCGCAGCAGAGATTCGTGACTCCCGAGA 364
    |||:|||||
265 rArgAsnSerCysSerHisSerProSerArgArgValTrpArg...LeuT 281
365 CGGAGCCAGAGAGAGGCGCATGAGCGACACACAAACAGAACCA 414
    |||:|||||
281 hTrpThrProGly.....SerThrProSerArg 289
415 ACAGCCAGTCCAGAGGCCCA 435
290 ProAlaSerAlaArgProPro 296
seq_name: sp.rodent:O63134

seq_documentation_block:
ID O63134 PRELIMINARY; PRT; 246 AA.
AC O63134;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE CONTIGUOUS REPEAT POLYPEPTIDE PRECURSOR.
GN CRP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=87165965; PubMed=3558393;
RA Heinrich G., Habener J.F.;
RT "Genes encoding proteins with homologous contiguous repeat sequences
RT are highly expressed in the serous cells of the rat submandibular
RT gland.";
RL J. Biol. Chem. 262:5262-5270(1987).
DR EMBL; M31032; AAA40971.1; -.

```

DR EMBL: M17703; AAA0971.1; JOINED.
 KW Signal. 1 18 POTENTIAL.
 RT SIGNAL. 19 246 CONTIGUOUS REPEAT POLYPEPTIDE.
 FT CHAIN 19 246
 SO SEQUENCE 246 AA; 26480 MW; 5BD4BBA0477A92DC CRC64;

alignment_scores:

Quality: 103.00 Length: 139
 Ratio: 1.304 Gaps: 7
 Percent Similarity: 56.835 Percent Identity: 28.058

alignment_block:

US-09-651-563-808 x Q63134 ..

Align seg 1/1 to: Q63134 from: 1 to: 246

```

306 GGGGTATCCGAGTCCGAGACGACCTGAACCCCGACAGAGATTCTGCA 355
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
74 GlySerGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 90
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
356 CTTCCCAAGCGGACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 405
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
90 ngLupProAlaThr.....SerGlyS 98
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
406 CAGAACCAACAGACGACGACGACGACGACGACGACGACGACGACG 446
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
98 erGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 114
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
447 CCC.....CAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 466
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
115 ProProAlaThrSerGlySerGlySerGlySerGlySerGlySerGly 131
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
467 A...GCTGAAGTCGGGATCTACACCTGGGCGAGAGAGAGAGAGATC 513
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
131 rGluAlaGluGluGluGluGluGluGluGluGluGluGluGluGluG 148
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
514 AGGATACAGCTGATCCGACGACGACGACGACGACGACGACGACGAC 563
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
148 lngGluGluGluGluGluGluGluGluGluGluGluGluGluGluG 160
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
564 CTGCATCACTCAAAACCGGGGATTAATCGATTGGTCCGGCGCTCA 613
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
161 .....ProSerAspSerAlaGlyGluGluGluGluGluGluGlu 172
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
614 AGGTGAAGTATTAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 663
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
172 nProGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 188
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
664 GAAGAGCAACACCAA 678
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
189 GlnGluGluGluGluGluGluGluGluGluGluGluGluGluGluG 193
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:

```

seq_name: sp_invertebrate:Q9U229

seq_documentation_block:

ID Q9U229 PRELIMINARY; PRT; 643 AA.

AC Q9U229:

DT 01-MAY-2000 (TREMblrel. 13, Created)

DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)

DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)

DE Y56A3A.32 PROTEIN.

GN Y56A3A.32.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;

OC Rhabditidae; Telodermidae; Caenorhabditis.

OX NCBI_TaxID=6239;

ON [1]

RP SEQUENCE FROM N.A.

RA Matthews L.;

RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=99069613; PubMed=9851916;

RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RT Investigating biology."
 RL Science 282:2012-2018(1998).
 CC -1- COFACTOR: FAD (BY SIMILARITY).
 CC -1- SIMILARITY: TO PYRIDINE NUCLEOTIDE-DISULPHIDE OXIDOREDUCTASES
 CC CLASS=1

DR EMBL: AL132860; CAB60511.1; -.
 DR InterPro: IPR001100; -.
 DR InterPro: IPR001327; -.
 DR Pfam: PF00070; Pyr_redox; 1.
 DR PRINTS: PR00368; FADPNR.
 DR PRINTS: PR00411; PNDRTASEI.
 KW FAD; Flavoprotein; Oxidoreductase; Redox-active center.
 SO SEQUENCE 643 AA; 71257 MW; EAC5A54980A5F75D CRC64;

alignment_scores:

Quality: 103.00 Length: 141
 Ratio: 1.272 Gaps: 8
 Percent Similarity: 57.447 Percent Identity: 26.241

alignment_block:

US-09-651-563-808 x Q9U229 ..

Align seg 1/1 to: Q9U229 from: 1 to: 643

```

207 GTGCCAGCCCA.....TGACCTTCTGTCGTCACGGCCA 244
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
45 IleProLysProGlySerLeuAspThrPhePheSerArg...SerH1 60
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
245 TAACTAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 294
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
60 sThrLysSerAlaHisGluPheGluProTyrLys.....ProGluLeu 75
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
295 GGGTGTGCTGGGG.....GTATCGAGTCCAGAGAGAGAGAGAGAGAG 332
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
75 lYalaPheIleGlyAlaValAlaPheIleGlyLeuThrLeuIleAlaVal 91
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
333 GAACCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 382
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
92 ValIleLysThrAspValPheLysLysGluAspSerHisLysGlyH1 108
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
383 GGCATGAGCCAGACACACA.....AACACAGAACACACAGCCATCC. 425
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
108 sGlyHisAlaLysHisSerLysLysHisGluGluLysHisGluLysH 125
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
426 ..CAGGAGCCAGTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 467
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
125 lsgLugluLysGluHisAlaGluProGluLysLysGluGluAlaLysPro 141
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
468 .....GCTGAAGTCGGGATCTACACCTGGGCGAG..... 497
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
142 GluLysProAlaGluProLysGluProGluProAlaGluLysGluAlaG 158
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
498 .....CAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 537
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
158 uGlnProGluGluAlaGluGluLysGluGluGluLysAspAlaGluPro 175
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
538 CGGACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 560
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
175 ysgLugluValaLysAspArgGln 182
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:

```

seq_name: sp_virus:Q9DUM3

seq_documentation_block:

ID Q9DUM3 PRELIMINARY; PRT; 1036 AA.

AC Q9DUM3:

DT 01-MAR-2001 (TREMblrel. 16, Created)

DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)

DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)

DE LATENT NUCLEAR ANTIGEN (FRAGMENT).

OS Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Rhadinovirus.
 RX NCBI_TaxID=37296;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Pirot T., Trimler M., Coppey M., Nicolas J.C., Marechal V.;
 RT "Close but distinct regions of LNAI are responsible for nuclear
 targeting and binding to human mitotic chromosomes.";
 RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF305694; AAC27458.1; -
 FT NON_TER 1036 AA; 119328 MW; 2959ED2C1C042B8 CRC64;
 SQ SEQUENCE 1036 AA; 119328 MW; 2959ED2C1C042B8 CRC64;

alignment_scores:

Quality: 102.50 Length: 143
 Ratio: 1.424 Gaps: 5
 Percent Similarity: 50.350 Percent Identity: 23.077

alignment_block:

US-09-651-563-808 x Q9DUM3 ..

Align seg 1/1 to: Q9DUM3 from: 1 to: 1036

```

261 AGGGCCGAGAGTGGAGGGGCTCAGGCGAAGCTGGTGTGGGGGT 310
    :: ||||| :: ||||| :: ||||| :: ||||| :: |||||
432 Lysylsglnaspqlnglnaspqlnglnaspqlnglnaspqlngln 448
    :: ||||| :: ||||| :: ||||| :: ||||| :: |||||
311 ATCCGAGTCCAGAGACCTCGGACCCCGACAGAAATCTGTGACTCC 360
    :: ||||| :: ||||| :: ||||| :: ||||| :: |||||
448 eglntserterglnglnnglnnglnnglnnglnnglnnglnngln 460
    :: ||||| :: ||||| :: ||||| :: ||||| :: |||||
361 CAGAGGGGACGAGAGAGGAGCGCATGAGCAGACACAAACAGAA 410
    :: ||||| :: ||||| :: ||||| :: ||||| :: |||||
461 .....Glu 461
411 CCACACAGC.....CAGTCCAGAGAGCCAGTAATGAGAGCCCA 451
    :: ||||| :: ||||| :: ||||| :: ||||| :: |||||
462 Proglnglnnglnnglnnglnnglnnglnnglnnglnnglnngln 478
    :: ||||| :: ||||| :: ||||| :: ||||| :: |||||
452 AAAAGAGAACGACAGCTGAAAGTCCGATCCTACACCTGGGAGAGA 501
    :: ||||| :: ||||| :: ||||| :: ||||| :: |||||
478 nglnglnnglnnglnnglnnglnnglnnglnnglnnglnngln 495
    :: ||||| :: ||||| :: ||||| :: ||||| :: |||||
502 CAGAGAGAGATCAGATCAGATCAGATCCAGTCCGATGAGAGGT 551
    :: ||||| :: ||||| :: ||||| :: ||||| :: |||||
495 lnpProglnglnnglnnglnnglnnglnnglnnglnnglnngln 505
    :: ||||| :: ||||| :: ||||| :: ||||| :: |||||
552 GATCTGAGAGAGCTGATCAGTCAACACCGGGGATTAATCTGATTGG 601
    :: ||||| :: ||||| :: ||||| :: ||||| :: |||||
506 Proleuglnnglnnglnnglnnglnnglnnglnnglnnglnngln 514
    :: ||||| :: ||||| :: ||||| :: ||||| :: |||||
602 GTTCCGGCGTCAAGTGAATATATACCTAAGAGAGACACTGAAATG 651
    :: ||||| :: ||||| :: ||||| :: ||||| :: |||||
515 .....Proglnglnnglnnglnnglnnglnnglnnglnngln 522
    :: ||||| :: ||||| :: ||||| :: ||||| :: |||||
652 CCAGAGCAGGTGAAGACACACCA 678
    :: ||||| :: ||||| :: ||||| :: ||||| :: |||||
523 GlnlupProglnglnnglnnglnngln 531
    :: ||||| :: ||||| :: ||||| :: ||||| :: |||||

```

seq_name: sp_virus:098148

seq_documentation_block:

ID 098148 PRELIMINARY: PRT: 1162 AA.

AC 098148:
 DT 01-FEB-1997 (TRENBLREL 02, Created)
 DT 01-FEB-1997 (TRENBLREL 02, Last sequence update)
 DT 01-JUN-2000 (TRENBLREL 14, Last annotation update)
 DE ORF73 HOMOLOG.
 OS Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Rhadinovirus.

OX NCBI_TaxID=37296;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97048116; PubMed=8892957;
 RA Cesarman E., Nador R.G., Bai F., Bohenzky R.A., Russo J.J.,
 RA Moore P.S., Chang Y., Knowles D.M.;
 RT "Kaposi's sarcoma-associated herpesvirus contains G protein-coupled
 RT receptor and cyclin D homologs which are expressed in Kaposi's sarcoma
 RT and malignant lymphoma.";
 RL J. Virol. 70:8218-8223(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97094384; PubMed=8939871;
 RA Moore P.S., Bashoff C., Weiss R.A., Chang Y.;
 RT "Molecular mimicry of human cytokine and cytokine response pathway
 RT genes by KSHV.";
 RL Science 274:1739-1744(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97121480; PubMed=8962146;
 RA Russo J.J., Bohenzky R.A., Chien M.C., Chen J., Yan M., Maddalena D.,
 RA Parry J.P., Peruzzi D., Edelman I.S., Chang Y., Moore P.S.;
 RT "Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus
 RT (HHV8).";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:14862-14867(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Russo J.J., Bohenzky R.A., Chien M.C., Chen J., Yan M., Maddalena D.,
 RA Parry J.P., Peruzzi D., Edelman I.S., Chang Y., Moore P.S.;
 RL Submitted (OCT-1996) to the EMBL/Genbank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Russo J.J., Bohenzky R.A., Chien M.C., Chen J., Yan M., Maddalena D.,
 RA Parry J.P., Peruzzi D., Edelman I.S., Chang Y., Moore P.S.;
 DR EMBL: U52064; AAC55944.1; -
 DR EMBL: U75698; AAC57158.1; -
 DR Interpro: IPR02017; -
 SQ SEQUENCE 1162 AA; 135213 MW; 1A72CE01C1CB081C CRC64;

alignment_scores:

Quality: 102.50 Length: 138
 Ratio: 1.235 Gaps: 5
 Percent Similarity: 52.532 Percent Identity: 22.785

alignment_block:

US-09-651-563-808 x Q98148 ..

Align seg 1/1 to: Q98148 from: 1 to: 1162

```

261 AGGGCCGAGAGTGGAGGGGCTCAGGCGAAGCTGGTGTGGGGGT 310
    :: ||||| :: ||||| :: ||||| :: ||||| :: |||||
422 Lysylsglnaspqlnglnaspqlnglnaspqlnglnaspqlngln 438
    :: ||||| :: ||||| :: ||||| :: ||||| :: |||||
311 ATCCGAGTCCAGAGACCTCGGACCCCGACAGAAATCTGTGACTCC 368
    :: ||||| :: ||||| :: ||||| :: ||||| :: |||||
438 eglntserterglnglnnglnnglnnglnnglnnglnnglnngln 455
    :: ||||| :: ||||| :: ||||| :: ||||| :: |||||
339 .....CGACAGAGATCTGACTCCCGACAGCGGAGCAGAGAGA 377
    :: ||||| :: ||||| :: ||||| :: ||||| :: |||||
455 lnglupProglnglnnglnnglnnglnnglnnglnnglnngln 467
    :: ||||| :: ||||| :: ||||| :: ||||| :: |||||
378 GGGAGCGATGAGCAGACACACACACACACACACAGCAGCAGTCC.. 425
    :: ||||| :: ||||| :: ||||| :: ||||| :: |||||
468 .....Glnnglnnglnnglnnglnnglnnglnnglnngln 476
    :: ||||| :: ||||| :: ||||| :: ||||| :: |||||
426 .....CAGAGCCAGTAATGAGAGAGCCCAAAAG 456
    :: ||||| :: ||||| :: ||||| :: ||||| :: |||||
476 oglnnglnnglnnglnnglnnglnnglnnglnnglnnglnngln 493
    :: ||||| :: ||||| :: ||||| :: ||||| :: |||||
457 AGACACACAGCTGAAAGTCCGATCCTACACCTGGGAGCAGACAGAA 506
    :: ||||| :: ||||| :: ||||| :: ||||| :: |||||

```


Fri Jun 29 10:17:50 2001

us-09-651-563-808.rspt

Page 11

